R E P O R T   R E S U M E S

ED 017 911 48 AL 001 112

TWO BEHAVIORAL EXPERIMENTS ON THE LOCATION OF THE SYLLABLE BEAT IN CONVEPSATIONAL AMERICAN ENGLISH.

BY- ALLEN, GEORGE DOUGLAS
MICHIGAN UNIV., ANN ARBOR,CTR.FOR RES.LANG.AND BEH

REPORT NUMBER BR-6-1784 PUB DATE 67
CONTRACT OEC-3-6-061784-0508
EDRS PRICE MF-$0.75 HC-$7.24 179P.

DESCRIPTORS- *SYLLABLES, *LANGUAGE RESEARCH, *ENGLISH, EXPERIMENTS, *BEHAVIORAL SCIENCE RESEARCH, SPECTROGRAMS, *STRESS VARIABLES, STANDARD SPOKEN USAGE,

RHYTHM IS ONE OF THE ELEMENTS OF THE PROSODIC LEVEL OF SPEECH. THE BASIC UNITS OF RHYTHM, UPON WHICH ARE CARRIED MANY OTHER PROSODIC UNITS, ARE POINTS OR INTERVALS OF TIME. THE TIME INTERVALS BETWEEN SUCCESSIVE MAJOR STRESSES ARE HYPOTHESIZED TO REMAIN ROUGHLY EQUAL IN ENGLISH SPEECH, I.E., ENGLISH IS SAID TO BE A STRESS-TIMED LANGUAGE. IN ORDER TO MEASURE THESE TIME INTERVALS, HOWEVER, THEIR END POINTS, THE RHYTHMIC BEATS, MUST FIRST BE FOUND. NATIVE SPEAKERS OF ENGLISH FEEL THE RHYTHM OF THEIR SPEECH INTUITIVELY AND CAN REACT CONSISTENTLY TO THE BEAT OF A STRESSED OR RHYTHMIC SYLLABLE. THE PRESENT WORK DETERMINED THE VALIDITY OF THIS RHYTHMIC INTUITION AND USED IT TO LOCATE THE SYLLABLE BEAT. OF THE TWO MEANS OF LOCATING THE SYLLABLE BEAT, (TAPPING TO THE BEAT WITH A FINGER, AND BY PLACING AN AUDIBLE CLICK ON THE BEAT), TAPPING SEEMED TO BE A MORE VALID RESPONSE. FROM THE FINDING OF AGREEMENT BETWEEN RHYTHMICALNESS-STRESS AND TAPPING BEHAVIOR, IT CAN BE CONCLUDED THAT RHYTHM EXISTS IN CONVERSATIONAL ENGLISH, INSOFAR AS THE STIMULUS UTTERANCES USED IN THIS EXPERIMENT ARE REPRESENTATIVE OF CONVERSATIONAL ENGLISH. THIS PAPER APPEARS IN "STUDIES IN LANGUAGE AND LANGUAGE BEHAVIOR, PROGRESS REPORT IV," PUBLISHED BY THE CENTER FOR RESEARCH ON LANGUAGE AND LANGUAGE BEHAVIOR, UNIVERSITY OF MICHIGAN, ANN ARBOR, MICHIGAN 48108. APPENDICES C AND D OF THIS REPORT (NOT INCLUDED) MAY BE OBTAINED BY WRITING DIRECTLY TO THE AUTHOR AT THE CENTER. (AUTHOR/AMM)

ED017911

TWO BEHAVIORAL EXPERIMENTS ON THE LOCATION OF THE SYLLABLE BEAT IN CONVERSATIONAL AMERICAN ENGLISH

By George Douglas Allen

Rhythm is one of the elements of the prosodic level of speech. The basic units of rhythm, upon which are carried many other prosodic units, are points or intervals of time. The time intervals between successive major stresses are hypothesized to remain roughly equal in English speech, i.e., English is said to be a stress-timed language. In order to measure these time intervals, however, their end points, the rhythmic beats, must first be found. Native speakers of English feel the rhythm of their speech intuitively and can react consistently to the beat of a stressed or rhythmic syllable. The present work determined the validity of this rhythmic intuition and used it to locate the syllable beat.

Previous investigators of speech rhythm have located the syllable beat by tapping to the beat with a finger and by placing an audible click on the beat. The present work studied the reliability and validity of these two behavioral tasks as measures of syllable beat location. The investigation of reliability calibrated the variability among and within subjects in reacting to syllable beats and identified different sources of variability. The validity studies were of two kinds: the first matched experimentally obtained differences in behavior with intuitively perceived differences in speech rhythm; the second abstracted from the experimental data rules for locating the syllable beat.

The reliabilities of the two tasks were found to be approximately the same: responses to stresses and rhythmic syllables showed a standard error of approximately three hundredths of a second for both tasks. Tapping seemed, however, to be a more valid response than placing a click, both for determining rhythms and for locating the syllable beat. The magnitude of subjects' variances in tapping to a syllable was found to correlate highly with: (1) the role of the syllable in the rhythm of the utterance, according to the experimenter's and the subjects' intuitions; (2) the stress markings assigned by linguists to the syllable; and (3) the grammatical class of the word to which the syllable belonged. Specifically, the syllables with lower tapping variances were felt by the experimenter and the subjects to be stressed or rhythmical syllables; they were more often marked as stressed by trained linguists; and they were more likely to be the lexically stressed syllables of open-class words, such as nouns, verbs, or adjectives.

The means of the distributions of responses were subject to biases (displacements) resulting from differences between subjects, differences between syllables, and differences within the subjects over time. These biases were more consistent and more easily calibrated in the tapping experiment than in the click placing experiment. The displacement of a subject's responses to a syllable was found to relate to the length of the consonant sequence preceding the nuclear vowel of the syllable.

Because agreement was found between perceived stress rhythm and tapping behavior, it was concluded that conversational English has rhythm and might therefore be stress-timed. Since bias in the location of tapping can be calibrated, the time between successive beats in the rhythm of an utterance can be measured; therefore the hypothesis that English is a stress-timed language may be tested.

CHAPTER I

Introduction and Survey of Pertinent Literature

1.0 Introduction

Rhythm is the structure of temporal intervals in a succession of events. An investigation of the rhythm of English necessarily encompasses not just the description of the temporal patterns of English, but also a description of the relationship of those patterns to the time-keeping abilities of English speakers and, in part, an understanding of the mechanism that produces and perceives the "events" of speech. Briefly, the term 'stress-timing' (or isochronism), as it is applied to English speech, means that the stresses, whatever they may be, mark off equal time periods. Here a "stress" is an event which we assume to occur at a point in time; thus, the temporal distance between successive "stress-points" remains constant. Physical time is a continuous and uniform phenomenon, as far as behavioral science is concerned, but behavioral time is something notoriously dis-continuous and non-uniform. The "constancy" of the temporal distance between events may therefore be considered as either physical or behavioral, with a corresponding change in meaning of the definition of "speech rhythm." For it is easily demonstrated that physically equal time periods can be perceived as quite different, and perceptually equal periods are measurably different (Woodrow, 1909, 1951). The other undefined word in the definition of isochronism is "stress". Stress is felt intuitively by native speakers of English, but its exact nature has eluded description. Various definitions have been offered in terms of intensity of the speech wave, pitch level, muscular or psychophysical effort of production, and other phenomena, along with all possible combinations. The treatment of this problem by Fonagy (1958) is good, although biased toward a definition in terms of muscular activity. It is obvious that different conceptions of stress will lead to different definitions of "stress-timing"; care must be taken to remain as free as possible from a priori constraints when answering an ill-defined problem. It seems clear that whatever stress may be for stress-timing, it must be perceived as a rhythmic accent. Native speakers feel that their speech is rhythmical, but this rhythm will be defined by accents whose accentual role extends to other functions, such as lexical stress and contrastive accent (Trager, 1941); there may be accents whose existence as accents derives solely from rhythmic constraints.

There is the further problem of attributing to a given "stress" a time of occurrence. For if temporal patterns are to be given anything more stable than an intuitive definition, actual time periods must be measured. These time periods require end-points, and for the measurement of isochronism, these points must be the times of occurrence of the stresses.[1]

1.1 Nature of the Thesis

It is the purpose of this research to investigate rigorously some behavioral aspects of isochronism in spoken English. "Rigor" implies that intuitive concepts are acceptable in theoretical statements, if, and only if, there is agreement between the intuitions of various native speakers. An example of a term not acceptable in this light

[1]It is quite possible that "point" is not the best term for describing the event in question (occurrence of stress), and that time periods must be described as statistical quantities. This shift of emphasis away from absolute time events is methodologically very useful in measuring behavioral time.

would be "stress", whose precise nature and function has so far eluded consensual definition. Whenever there is agreement, however, it will serve as a fact from which to build toward a theory. The choice of the behavioral domain is based on a belief that not enough is understood about the mechanism of speech production, the structure of language, or their interaction to warrant conclusions about speech rhythm. Rhythm is a basic element of prosody, and prosody is that level of language where the message and the process of expression interact most closely.

The purpose of Chapter I is to review the historical matter pertinent to speech rhythm so that a reasonable statement of the meaning of stress-timing can be made. The five general categories into which the literature has been subdivided in Section 1.2 correspond to different behavioral aspects of the problem. Sub-sections 1.2.1 and 1.2.2 concern themselves with the perception of time and rhythm, respectively. Since behavioral time is not identical with physical time, it is important to know the relationships between the two. The manner in which rhythm is perceived is important since, if speech rhythm exists, it probably is as important for the listener as it is for the speaker. Kinesthesis, the bridge between production and perception in the motor domain, is treated in Section 1.2.3. The last two sections review work on motor production of syllables and stress (1.2.4.) and behavioral studies attempting to locate a rhythmic beat associated with syllables or stresses (1.2.5.). Following the review is a recapitulation of Section 1.2 (1.3.1), a discussion of how one might prove the existence of stress-timing in English (1.3.2), and a statement of the purpose of the present study (1.3.3).

The empirical studies reported in Chapters II and III represent an attempt both to exhibit behavioral regularities in the perception of rhythm and, more importantly, to make very clear what kinds of statements may be induced from the data. The form of the experiments parallels to some extent previous studies on rhythm in English (Miyàke, 1902; Hollister, 1937; Classe, 1939; Newcomb, 1961), and some comparison with this historical matter is therefore possible. Data analysis is carried further, however, than in these previous studies, with a resultant increase in confidence in making probabilistic statements. Although the development of the techniques of experimental design has been considerable in the last few decades (post- R. A. Fisher), recent investigators have either ignored or been unaware of the existence and utility of these techniques.

1.2 Survey of Pertinent Literature

Experimental studies of rhythm and time perception (two closely related phenomena) are too numerous to be reviewed here completely. This discussion concerns only the results that bear on the problem of speech rhythm.

1.2.1 Perception of Time

Long intervals tend to be perceived and reproduced as shorter than they actually are, and short intervals tend to be perceived as longer. This centralizing tendency has resulted in experimental definition of the so-called "indifference interval" (Woodrow, 1951, p. 1225) which is neither shortened nor lengthened. Investigators have found differing values, depending on their subjects and research methods, but a value of 0.5 to 0.7 sec, seems to have some generality (Woodrow, 1951, Fraisse, 1963).

Fraisse (1963) reviews extensively the work on the auditory perception of duration, in which the successive defining events are relatively simple sounds, such as clicks or pure tones. He found that for two successive temporal intervals, defined by four acoustic events (where the 2nd and 3rd may in fact be the same event, yielding two immediately successive durations), the perceptual judgment of the relative duration of these two intervals depends on many different factors. The more important of these for the present work are the intensity, pitch, and duration of the delimiting stimuli (Fraisse, 1963; Woodrow, 1909). For example, an interval seems shorter in duration as the two stimuli bounding it become more intense, provided these two stimuli are of equal intensity, or if the more intense stimulus begins the interval; however, for short intervals, if the final stimulus is the more intense, greater intensity leads to shorter perceived duration. With respect to pitch, the higher the pitch of the delimiting stimuli, the longer will seem the interval. Also, the greater the difference in pitch between the first and second sounds, the greater will be the perceived duration, as long as there is little harmonic relation between the pitches (such as an octave difference). If the duration of the delimiting sounds, or just that of the first one, is increased, the perceived duration increases; if the longer stimulus comes last, the interval seems shorter.

All of the above results refer to so-called "empty time" intervals, that is, intervals set off by two stimuli between which no physical signal is present. However, there is very little silence during the speech process; it is necessary, therefore, to know what happens to the perception of an interval when it is "filled". Fraisse reports that the indifference effect also holds for filled intervals, that is, shorter (longer) intervals are over-(under-)estimated where shorter (longer) is with respect to the "indifference interval", presumably of the same order of magnitude as for unfilled time. If an interval is subdivided by discontinuous stimuli," ... a divided interval appears longer than an empty interval of the same duration ... and an interval with more divisions appears longer than one with fewer ... Furthermore, the (interval) that is evenly divided appears longer than that which is irregularly divided" (Fraisse, 1963, pp. 132-133). More intense or higher pitched continuous sounds will appear longer than less intense or lower pitched ones. Although the intensity levels used in the various studies are not reported, the pitch values range from just over 100 cps to 3000 cps, a valid range for speech studies.

The results of studies comparing empty with filled intervals are not clearcut, according to Fraisse (1963). If the empty interval follows the filled, the first filled interval appears longer (provided they are physically equal); however, this effect is open to interpretation on a different ground, namely that of the attitude of the listener. Fraisse suggests that the change in perception may be due to a change in the listener's focus of attention, this process of focusing being essentially different from the perceptual processes governing the other temporal judgments studied. Other results on filled vs. unfilled intervals show little or no difference in perception or behavior toward them.

When subjects are asked to reproduce a standard interval by some behavioral response, usually tapping the finger, the accuracy of reproduction is different for different sized intervals, (Woodrow, 1951). Accuracy seems greatest at the lower end of the interval size scale, i.e., from 0.2 sec to 2.0 sec, where the standard error of reproduction (i.e., the standard deviation of the distribution of reproductions) is about eight per cent of the given interval. The standard error increases to about 16 per cent in the 4 to 30 sec interval-size range.

But for purposes of speech rhythm studies, the lower range would seem more appropriate, since the rate of temporal succession of rhythmical stresses usually falls within those bounds.

In the comparison of two given intervals, attitude has an effect on perception, as was suggested in the comparison of empty with filled intervals. Woodrow (1951) suggests that two different oppositions of attitudes are in evidence in existing data (see also Fraisse, 1963). The first opposition is that of the subject's perception of the two intervals plus the intervening pause as a single unitary pattern, with whatever accents and rhythms may result, as opposed to his perception of the second interval as a stimulus to be matched against the first one, for which he has an immediate memory. In the first case, judgments would presumably be based on the overall effect of the two intervals plus the intervening pause; in the second case, judgments would probably be based on the results of the matching process. The other opposition is that of objective to subjective attention to the stimuli. "In the objective attitude attention is centered upon characteristics of the stimulus ...; in the subjective attitude the subject intentionally abstracts from, or ignores the objective stimulus and concentrates upon the experience of duration" (Woodrow, 1951, p. 1128). Woodrow takes care to point out that these oppositions are meaningless unless some behavioral differences can be shown to arise from them through "differential instruction" of subjects. He cites two examples in which (1) subjects' perceptions of interval length changed as a result of active vs. passive attention to the second interval and (2) the length of the reproduced interval changed with attention to the stimuli and to the duration.

1.2.2. Perception of Rhythm

Another set of studies has concerned itself with the way in which rhythm is perceived, that is, the manner in which stimuli become organized temporally into a structured pattern. *"The nature of the perceived grouping ... of a series of stimuli ... is largely, but not entirely determined by the characteristics of the stimulus series. The most important of these characteristics are the relative intensities of the members of the series, their durations, both absolute and relative, and their temporal spacing ... For example, with equal temporal spacing, and not too fast a rate, and every second sound louder than the others, the series of sounds tends to be heard in groups of two, with the louder sound beginning the group. If, however, the interval following the softer sound is decreased while that before it is correspondingly increased, a point is reached where grouping occurs with the softer first and the louder sound second. ... As regards the effect of the relative duration of the stimulus, when intensity and temporal spacing are uniform, if every second sound is longer, the probabilities are in favor of ... the second sound [as] the second member of the group."* (Woodrow, 1951, p. 1223).

Even if there is no objective difference between the stimuli in the sequence, rhythmic grouping still takes place. This grouping can be a result of subjective, involuntary, kinesthetic movement, of which more will be said in the next section, or of intentional movements or other rhythmic actions. In the absence of objective difference, any rhythmic grouping must be entirely subjective. "The number of members grouped together in one rhythmical measure is increased from two to six, or more, with increase in rate." (Woodrow, 1951, p. 1233). Different rhythmic groupings can take on different meters (accent patterns).

The important point here is that rhythmic grouping apparently is natural and perhaps unavoidable in the perception of temporal phenomena. This would suggest that a person looking for rhythm in the temporal succession of stresses would likely perceive it subjectively in the absence of an objective rhythm.

1.2.3 Kinesthesis

One of the fascinating aspects of rhythm is the kinesthetic reactions we have to it. The seeming inexorability with which motor reactions accompany our perception of rhythm has forced many investigators to the conclusion that rhythm, when viewed as a behavioral phenomenon, is largely a motor activity. This means that the organization or "structuring" of the temporal sequence is carried out in the motor system rather than, say, the auditory or higher-order language system. If this is so, then a lot can be said about the role of rhythm in speech, for one can focus attention on the purely motor aspects of speech for the definition of this role. Stetson, like many workers before him, assumed that rhythm of the kind we have talked about above is basically a motor phenomenon. He wrote that "... as a general theory, the motor hypothesis needs no defense. Its only competitor was the 'mental activity' theory which is manifestly incapable of explaining the peculiarities of the unit-groups and of larger groupings" (1905, p. 225). Stetson did not deny that rhythm can have affective or emotional meaning but he did believe that such emotional coloring results from an interaction of the motor system with these other systems. He believed that the perception and interpretation of rhythm are always mediated through the motor system. He wrote at length on the different muscular movements and strains that accompany the perception of rhythm. Ruchmich (1913) studied kinesthesis in the perception of rhythm, attempting to find from introspective reports the different kinds of kinesthesis and their importance in the perception of rhythm. He concluded that kinesthesis changes qualitatively as well as quantitatively throughout long attention to a rhythmic stimulus, but more importantly, he found that visual imagery or auditory imagery or sensation can substitute for kinesthesis in rhythm perception <u>only after the perception was established</u>. In other words, kinesthesis was necessary for the primary perception of rhythm. This primary role becomes all the more important for the study of speech rhythm when we discover the degree to which the speech mechanism is active in this kinesthesis. Stetson wrote "... the most important natural rhythm-producing apparatus is the vocal apparatus... The tongue... and the muscles of respiration play a frequent part in rhythmicization" (1905, p. 257). Woodrow elaborated on the sensations of strain that accompany the subjective attitude toward rhythm as follows: "The strain may be described as a feeling, as strain of attention, as strain of expectation or as a group of kinesthetic sensations. The sensation of strain may apparently originate in almost any part of the body. Frequently mentioned are sensations of strain from the arms or hands, from the muscles involved in breathing, and from the vocal organs, the latter being sometimes accompanied by auditory imagery" (1951, p. 1228). Bolton reported the most common movements to be of the foot, head or trunk, but also mentioned that "... slight or nascent muscular contractions were felt in the root of the tongue or larynx" (1894, p. 91).

1.2.4 Motor Activity in Rhythm and Stress

It is obvious that there is also muscular activity in the production of speech, but it has not been proven that muscular activity is the exclusive definer of speech rhythm. Since the syllable and its associated stress have always been central to the notion of rhythm in speech, the nature of these elusive units has been studied in relation to the muscular activities that might produce them. Considerable effort has been directed at describing the role of the respiratory muscles in "creating" syllables and stresses. Probably the foremost exponent of the respiratory-motor theory of syllable and stress production has been R. H. Stetson (see especially his Motor Phonetics, 1951), who believed that each syllable is initiated by a "pulse" of internal intercostal muscle activity, and that the main stresses of a breath group are produced by pulses of the abdominal muscles, principally the rectus abdominis. His conclusions were based on evidence drawn from pneumographic studies of body wall movements and some electromyographic data, both correlated with the actual utterance. Stetson's work has come under considerable criticism, however, and Cooker wrote that " ... respiratory activity during the production of speech is not well understood" (1963, p. 1). Several investigators have used electromyographic techniques in an attempt to identify specific muscle group activity during speech. Draper, Ladefoged, and Whitteridge, in a series of articles (Draper et al., 1957, 1959, 1960; Ladefoged et al., 1958; Ladefoged, 1960), proposed a fairly simple model for respiratory muscle activity in speech. They indicated that the respiratory muscles are used to maintain a relatively constant mean subglottal pressure as a source of air flow for the articulation of the speech at the larynx and above. They, like Stetson, compared the chest to a bellows, whose open end is the glottis and whose contracting force is supplied by the respiratory muscles. But while Stetson assumed that other "larger outside muscles and abdominal muscles" (1951, p. 1) account for the overall decrease in lung volume, Draper et al. (1959) imagined between the handles of the bellows a spring that produces a force analogous to the relaxation pressure of the lungs. This relaxation pressure is not constant, but varies directly with the volume of air in the lungs. Thus, in order to maintain relatively constant pressure below the glottis (and therefore in the lungs) throughout a long utterance, force must be applied against the relaxation pressure when, at high volumes, it is greater than what is needed and along with it when there is less air in the lungs. They interpret their electromyograms of various muscle groups as showing that the external intercostals are inspiratory in function and are used at the beginning of an utterance when the relaxation pressure is too high. The internal intercostals are expiratory, and are used more as lung volume decreases. They also are used in a pulsing manner to produce the stresses of the utterance. At very low volumes, other muscle groups assist the internal intercostals in further diminishing the chest cavity, these being the latissimus dorsi, rectus abdominis, internal and external obliques and the diaphragm. They did not find a separate pulse of intercostal activity for each syllable, but they did find a suggestive correlation between such activity and the major stresses of the utterance.

Eblen, as reported by Cooker, in a surface-electromyographic study of speech breathing, " ... suggested that muscle activity patterns during speech are strongly influenced by the depth of inspiration which precedes the utterance and closely related to the maintenance of a constant mean subglottal pressure" (Cooker, 1963, p. 12).

Fonagy investigated internal intercostal activity in stressed and unstressed syllables in short Hungarian, Russian, and English utterances using both surface and needle electrodes. In the study using surface electrodes he found that "... the muscle activity increased in every case on the accented syllable" (1958, p. 28). The use of needle electrodes produced a very high, but not perfect correlation, between relative stress of a syllable and the increase of internal intercostal activity (1958, p. 29, Table III).

Hoshiko (1962) found a different pattern of activity of the various respiratory muscles in speech from that pictured by Draper et al., although this result may have been due to differences in speech materials used. He used sponge surface electrodes, placed according to Stetson's and others' suggestions, to investigate the pattern of activity in the internal and external intercostals and the rectus abdominis. He found that all three muscle groups were active synchronously in the production of sequences of syllables, and that their pattern of activity did not change according to the rate of syllable production. He found that among the three muscle groups, the internal intercostals contract first, then the rectus abdominis, and finally the external intercostals. There was also a characteristic pattern of termination of muscle activity after the onset of phonation. Hoshiko found also that the internal and external intercostals work together in inspiratory as well as in expiratory phonation. These findings are in direct conflict with the model of strictly inspiratory and expiratory roles for the external and internal intercostal muscles, respectively.

Cooker noted both the difficulty with which electromyograms may be interpreted as showing activity of specific muscle groups and the amount of disagreement among investigators in the field. He studied the relationship between body wall movements, as measured by strain gages, and intra-oral pressure, in order to determine if their time relationships indicate that increases in pressure can be attributed to increased muscle activity or, conversely, the movements of the chest wall are attributable to back pressure from articulatory closure. He found that at slow rates of syllable repetition (one per sec) "... the sequential relationships ... indicate the presence of some type of preceding activity, possibly an expiratory contraction of the respiratory muscles, which is an integral part of the syllable..." (1963, p. 42). At higher rates of repetition (two and six per sec), movements of the chest wall were attributable to articulatory back-pressure in the cases of syllables with a stop consonant (/pa/ and /ba/), but not those with a laryngeal fricative (/ha/), where again the chest movement preceded the consonantal articulation. In the records for connected speech, Cooker found a closer correlation between chest wall movements and the number of vocal tract constrictions than he did for movements and syllables. He concludes his study with the inference that "... the speech breathing processes combine articulatory valving of a relatively steady background pressure with expiratory contractions of the muscles of respiration to produce the wide variety of pressures necessary for speech" (1963, p. 68).

Peterson (1958) found that a woman whose respiratory muscles were paralyzed but who had normal use of the laryngeal and supra-laryngeal articulatory musculature was able to produce "normal" speech during the expiratory phase of her iron lung. She appeared unable, however, to produce strong stress or loud speech. This evidence supports the theory that normal speech is, as Draper et al., and Cooker propose, a process of articulatory valving with the degree of stress correlated with the activity of the respiratory muscles.

1.2.5 Location of the Rhythmic Beat

The importance of motor systems in the rhythm of speech has received support from two areas of study, that on kinesthetic reactions to rhythmic stimuli and that on the functions of the respiratory and articulatory musculatures in the production of syllables and stresses, the obvious candidates for carriers of speech rhythm. Another area of investigation important for the elucidation of speech rhythm involves behavioral studies defining the "point of occurrence" of the stress or rhythmic beat by some motor activity, generally tapping of the finger or hand "in time" to the produced or perceived rhythmic stimulus, speech or otherwise. Miyake investigated the tapping behavior of subjects under various conditions of rhythmic constraint. In one study, he instructed subjects to move a lever, connected to a recording drum by means of a Marey tambour," ... up and down successively at irregular intervals at a rather rapid rate" (1902, p. 1). In a similar study subjects tapped a key in electrical connection with a Deprez marker and recording drum "... at intervals as irregular as possible, the slowest speed of two successive beats being limited to about one second" (1902, p. 2). He found repetitions of equal muscular intensity and alternation of strong and weak intensities in the first study, use of simple multiples of time intervals and alternation of long and short intervals in the second, and many successions of equal intervals in both studies. Further, the effort involved in behaving "arhythmically" was considerable.

In a second experiment, subjects beat on a noiseless key in time with either an auditory click or visual flash presented at a rate of one per second. The taps varied less in the auditory case than in the visual case, but there was marked tendency for the taps to precede the clicks (approximately five csec, on the average), this tendency not being as apparent in the visual case. Experiments of this kind for the auditory case have also been run by Johnson (1898) and Scripture (1397). Johnson found that subjects' taps preceded the click at first, but some subjects learned through practice to match the click more closely. The rate of click presentation was again one per second. Scripture (1899) noted that " ... most persons regularly beat time just before the signal occurs ...", but the data to which he referred (Scripture, 1897, p. 182) point to great variation between subjects and as much tapping after the beat as before.

Paillard (1946-1947) investigated the degree to which different motor organs (left and right index fingers, left and right heels and the lower jaw) could move simultaneously. The degree of precession of one organ by another was recorded by the deflection of an oscilloscope beam by the signal from a Wheatstone bridge whose two arms contained the contacts for the two organs. Under one condition subjects were told to move a given pair of organs simultaneously. Under the second condition, subjects moved them as soon as possible after an external signal was given. The resulting time differences for the various pairs of organs show these two tasks to be very different. In the first case, when the movement is voluntary, the organ moved first is the one which is farthest from the central nervous system in nerve transmission time. Paillard suggested that in this case the subject moves the two organs in such a way that the kinesthetic feedback impulses from the motions arrive in the cortex simultaneously. In the second case, where subjects reacted simultaneously, the time differences for the different organs seemed to be related to the simple reaction times for those organs. The order of magnitude of the non-simultaneity of motions was least for symmetric organs (right and left index fingers, or right and left heels) and less in the voluntary case then in the reactive case. A typical value would be one csec.

Miyake (1902) investigated the effect of auditory and visual feedback on tapping behavior. He constructed an electrical circuit so that the pressing of a noiseless key produced a spark whose visual and auditory energies were great enough to serve as feedback stimuli to mark the temporal occurrence of the tap. When the spark was behind a curtain, the subject could hear but not see it; the visual stimulus without the sound was effected by placing the spark gap inside the inner of two concentric glass tubes. Subjects tapped on the noiseless key at what they considered to be constant intervals of time, the rate of taps being left to their own choosing. Miyake found that the presence of an auditory feedback influenced subjects to produce generally shorter and more regular intervals. That is, both the mean interval length and the variability of these lengths about their mean decreased if the subject heard a click when he tapped. Visual feedback also tended to lower the mean interval length, but variability was not decreased. It should be noted that variability was reported as the ratio of the probable error ($= 2\sigma/3$, where σ is the standard deviation) to the average interval length, where there were usually ten intervals measured. There were 24 such ten-interval sets for four subjects in the auditory feedback task, and 28 ten-interval sets for three subjects in the visual feedback case. In a third series of experiments, Miyake investigated the effect of accentuation on produced interval size when subjects tapped and beat on a drum in various simple meters. He found, in agreement with Woodrow's (1909) perceptual experiments, that the interval whose beginning beat is accentuated is lengthened, and that the last interval in the metric group is lengthened. (Stetson [1905, ff.] showed similar results for iambic grouping, but his results for trochaic grouping were not clear.) Miyake ran two final series of experiments involving intensity and time in rhythmic speech. In both of these studies he used a laryngeal frequency transducer with a platinum diaphragm and the noiseless tapping key. Subjects' vocalizations caused the diaphragm to vibrate into contact with a platinum point, thus completing an electrical circuit. The times of contact and no contact were recorded on a revolving drum. Accurate measurements of laryngeal frequency, but not amplitude, were possible. The first of the two studies concerned itself primarily with the relative lengths of intervals demarcated by rhythmic phonations of the vowel /a/ as measured from the onset of phonation, and with the lengths of the phonations themselves, under different metrical conditions. Secondarily, Miyake studied the pitch relationships of the various accented and unaccented syllables. The interval lengths paralleled the results of the tapping experiment. Higher pitches and greater pitch changes were found to be associated with the accented syllable. Perhaps, as Miyake suggested, this correlation completes the tripartite interrelationship of greater intensity, longer duration, and higher pitch as accenting phenomena (see Section 1.2.1). The final experiment is of fundamental importance for the rhythm of speech. Subjects "... recited a syllable in a scanning manner while... beat[ing] time on the noiseless key with the finger of ... [the] right hand (generally the index finger), the rate of movement being left to [their] choice" (1902, p. 40). Nine syllables, /a/, /ʔa/, /ma/, /ha/, /pa/, /ap/, /a.p/, /mam/, and /ma·m/ were used. Each subject generated ten syllable-taps in a row for a given experimental trial. Not all subjects tapped to all the different phonetic syllable types. In each of the ten syllable-plus-tap records Miyake measured in msec the time interval by which the tap preceded the onset of the vowel. Again the mean and probable error ($2\sigma/3$) are tabulated along with the number of measurements used in each calculation and the numbers of cases in which the tap preceded and succeeded the vowel onset. The numerical values are given in Table 1. They show that the tap generally preceded the onset of the vowel. These data will be discussed further in relation to the data gathered in the present study. Miyake noted that one problem in interpreting these data is whether the tap can be taken as the actual "point in time" of the syllable beat, or whether errors caused by the time of neurological and mechanical transmission displace the tap from the perceived beat. He qualified his conclusions about the answer to that question. Note that subjects had been found earlier by Miyake and Johnson to tap before a perceived acoustic rhythmic stimulus, but this says nothing about where a tap would fall in relation to a self-generated rhythmic speech signal. Also, it is not clear that the onset of the vowel can be measured reliably in msec. Since the time of onset is taken as the time of the first contact of the diaphragm with the platinum point, weak vibrations or heavily attenuated vibrations would not be recorded. This objection applies less to the syllables /ʔa/ and /pa/, where the onset would be sharp, than to the others which begin with the vowel /a/, the voiced consonant /m/, or the laryngeal fricative /h/.

Wallin had subjects read pieces of poetry in such a way that the feet were equal in length. He measured the variability of the resulting interval lengths when the subjects tapped while they read and when they

Table 1

Summary of Miyake's Data for Tapping to Syllables

(After Miyake, 1902, p. 44)

Syllable	Average Time of Beat Before Vowel	Number of Measurements
/ma/	132 msec	210
/pa/	143	206
/ha/	118	190
/ʔa/	131	70
/a/	52	120
/ăp/	59	100
/āp/	52	90
/măm/	57	80
/mām/	62	80

did not, and found that the variability (probable error, as before) was less (one to two csec) when they tapped than when they did not (six to seven csec). (See Wallin, 1901, Tables LXII, p. 115, and LXXV, p. 121.) His method of interval measurement is open to criticism, however, as he slowed down the speech " ... until a group of rapidly recurrent sounds, which had previously appeared as a homogeneity, was split up into discrete elements of sound separated by gaps... It is the time of these several sensations [of sounds], as well as the time of the intervals of silence between them, that was measured" (1901, p. 24). What these sounds and silences correspond to in the acoustic signal is not clear, although he equated them with vibrations, both strong (sound) and weak (silence), and "changes in the condition of the vocal organs" (1901, p. 24). He does not state what relation these vibrations and conditions have to the rhythm, except that he wished to break the utterance into "separate syllable groups." Furthermore, data given in csec derived from measurements as gross as 1/32 of a second are not reliable.

Hollister (1937) investigated the movements of subjects' hands as they tapped while reciting poetry from memory; he wished to elucidate the nature of the "syllable impulse" as a motor-phonetic unit and felt that the hand movements would mirror subjective impulse feelings. He found that when subjects were instructed only to tap while reciting they tended to tap once for each syllable. Simultaneous kymographic recordings were made of subjects' vocalizations, by means of goldbeater's skin drum heads, extra-oral and extra-nasal pressure, by means of a flexible rubber dam connected pneumatically to the kymograph, and taps on two tambour boxes, one for the finger tips, the other for the heel of the hand, and covered with heavy rubber, also connected pneumatically to the kymograph. The utterance was also recorded by dictaphone.

Subjects made several records in various postures and under various instructions as to whether and how they should tap. Besides finding that subjects tend to tap once for each syllable, he found that the tap is synchronized with "...that point in a syllabic unit where the energies of breath pressure, vocalization, and articulatory stroke combined, reach their moment of climax... The average variation from simultaneity... is plus or minus .020 sec" (1937, pp. 82-83). Tapping did not seem to disturb the rhythmic patterns of speech that were produced in the absence of tapping. The hand tapping movements tended to match the syllable character in "duration of hold" and "intensity of hit" while passive and unconscious hand pressures similarly matched the longer phrases of the utterance. Unfortunately, he did not calibrate his apparatus for time delays in the different recording systems, and he gives no indication of the error measurement resulting from changes in recording speed of the kymograph. The speed was 61 mm per sec and his measurements were in mm, or, to the nearest 1/61 of a sec. Again, such an accuracy as in the .020 secs reported above is not justified by such a system.

Classe (1939) replicated Miyake's last experiment in an attempt to make a more general statement about the location of the syllable beat in rhythmic speech. He had subjects tap on a key while they recited isolated lines of verse, the different lines chosen for an overall representation of sequences of initial consonants on the accented syllables. The speech wave, the subjects' taps, and a time signal (frequency not given) were recorded on a smoked paper drum and measurements were made in units of .005 inch from the time of the subjects' taps to the time of explosion for plosive consonants and to the time of the onset of the nuclear vowel after all other consonants. He computed means and standard errors for these time differences and calculated confidence intervals of the mean equal to ±2 standard errors for the different consonant types. Using these confidence intervals, the largest of which was about 3 csec long, and shifting them according to a one csec lag time associated with the recording apparatus, he concluded that the "point" of stress (see Classe, 1939, p. 17) comes "(a) at the moment of plosion of breathed occlusives; (b) after the explosion of voiced occlusives; (c) at the moment of maximum deviation of the recording-pen for fricatives; (d) just before the beginning of the vowel in the case of all other consonants, except (h)" (1939, p. 45). He concluded further that this kind of stress is "... <u>mainly</u> a subjective notion... depend[ing] more on motor factors than on auditory ones." He finds his data to be in line with Miyake's, except in the case of /mam/. This alleged correspondence is difficult to see, at least in terms of absolute time measures. Miyake's data for the syllable /ha/ show the tap to precede the vowel by an average of .12 sec, while the corresponding figure from Classe's work is .004 sec. The 100 msec difference is not close agreement. For the syllables beginning with /m/ there is similar disagreement. No other comparisons between the two experiments are possible. The only agreement between the studies is in the conclusions, namely that the tap precedes the nuclear vowel in most syllables. Classe, like Miyake, thought the question of whether or not the tap can actually match the stress point to be important and unanswered (1939). Unlike Hollister, Classe thought the overall task of tapping to the syllable beat to be a difficult one (1963). As in previous studies, the degree of accuracy assumed was not justified. With a probable timing signal frequency of

100 cps, his unit of measurement of .005 sec required dividing each cycle of the timing signal into two. This is quite reasonable (see Classe, 1939, Figure 13, p. 33, for scale size) but his means and standard deviations could not then be given in msec, with three significant figures. Likewise he assumed his one csec delay factor to be valid for all recording conditions, and he admitted that only one significant figure is allowable in this factor. Another statistically questionable procedure was the pooling of data from taps gathered on different syllables, as in "sort" and "sinking" (1963, p. 26). It has not been established that the interval between tap and acoustic cue is not dependent on nuclear vowel or rate of speech; even the pooling of data for different utterances of the <u>same</u> syllable raises questions of the forms of the resulting distributions.

A different sort of experiment on location of syllable beats has yielded similar results. Newcomb (1960, 1961) wished to measure the change in syllable lengths under different conditions of rhythm as indicated by changes in the grammatical structure of utterances. In order to establish syllable <u>boundaries</u>, he had subjects synchronize their utterance of the successive syllables of a sentence with a sequence of clicks set at the rate of three or four pulses per second. Subjects apparently had no difficulty in this task, and Newcomb found by spectrographic analysis of the resulting speech plus clicks that "... the time marks coincide with the release of the last consonant before the onset of voicing" (1960, p. 29). In a second experiment, using a magnetized razor blade, he placed pulses by hand both at these "desirable" places and elsewhere in the syllable. "If these pulses were placed at any point other than the last consonant release they seemed to lose all relationship with the speech and became merely random noises" (1960, p. 31). He prepared pairs of sentences with clicks, one member of the pair having the clicks at the "posited perceptual syllable boundary", the other member marked "elsewhere". "The trained listeners ... had no trouble identifying the sentence <u>in which the pulses seemed to coincide with the rhythm of the syllables</u> [present writer's emphasis]" (1960, p. 32). In a later experiment he allowed subjects to move a click around in a given syllable of an utterance on a tape loop. Through hearing the utterance repeatedly, the subject was able to place the click at the location of "optimum sensation of coincidence [with the syllable]" (1961, p. 3). Again the preferred location was the point of "... release of the last consonant before the syllable nucleus... In the case of voiceless obstruent consonants, the point of demarcation occurs at the release of the consonantal articulation. When semivowels separate syllables, the point... falls at the beginning of the return from the extreme point of formant deflection toward the position of the following vowel." Although Newcomb was searching for syllable <u>boundaries</u>, it seems quite likely that subjects considered the click as an accent (or rhythmic beat, in the earlier case of repeated clicks) with which they were to match the syllable accent. This is especially clear in the experiment where subjects distinguished the sentence whose clicks matched the syllable rhythm from the one whose clicks did not. The fair agreement between the preferred click location and the times of tapping found by Miyake and Classe is further substantiation of this hypothesis. The disagreement in exact location may be attributable to the perceptual nature of Newcomb's experiments or to errors in the measurement process (his unit of measurement was one csec, and the distances between successive articulations can be of this order of magnitude).

1.3 Recapitulation, Discussion, and Statement of Purpose

1.3.1 Recapitulation of Section 1.2

Classe writes that absolute equality of consecutive time intervals is the exceptional case in English speech, even in poetry (1939). There must be metric, grammatical, and phonetic similarity between the speech segments comprising the two intervals before equality results. That such strict requirements are seldom met in conversation, however, does not rule out the existence of some form of speech rhythm. The studies on the perception of duration and on subjective rhythm, discussed in Sections 1.2.1 and 1.2.2 above, show that the listener is capable of expanding and contracting time so as to perceive a pattern that is not in the objective stimulus sequence. Logically prior to the perception of possible speech rhythms is the production of such rhythms. Section 1.2.4 concerns itself with some mechanisms for imposing rhythmic constraints on normal speech; the mechanisms mentioned are independent of any perceptual process. The link between the production and perception of rhythm in speech is kinesthesis, discussed in Section 1.2.3.

Sections 1.2.2, 1.2.3 and 1.2.4 can be put together to make a very weak, but appealing case for the existence of a relevant temporal structuring in normal speech. Miyake (1902) showed that it is difficult to carry out certain motor activities in anything but a rhythmic fashion. This may also be true of the motor activities of speech; but it is not proven. Thus, there is with some probability a temporal organization in the production of normal speech. Perceptually, moreover, listeners tend to impose subjective rhythms on stimulus sequences, regardless of the existence of any objective structuring in the sequence. They might do the same with speech. In order for a temporal structuring to have relevance for the communication process, however, there must be a close relationship between the rhythm the speaker produces and the rhythm the listener perceives. Studies on kinesthesis suggest the possibility of such a relationship. The argument hinges on the reports by subjects of kinesthetic tensions in the muscles of the chest, throat and larynx. It may be accidental that these speech organs are reportedly involved in kinesthesis, or it may be that they play some important part in our perception of rhythm. It is suggestive, however, that muscles that are involved in kinesthesis are perhaps the very same muscles that are used in speech production. The implication is that the produced rhythm and the perceived rhythm of speech, if they exist, are related by the common proprioceptive mechanisms of speech production and kinesthesis.

1.3.2 Discussion

As mentioned above, absolute equality of time periods is not to be expected in English speech, but there are techniques of detecting tendencies toward equality. One way in which a tendency toward equality could be shown is statistical in nature. The hypothesis of stress-timing implies that the time of occurrence of a stress is to some degree dependent upon the time of occurrence of the previous two stresses. That the timing of successive stresses is dependent upon the number of intervening syllables has been suggested by Classe (1939) and Halliday (1963) both of whom made measurements in support of this suggestion. A statistical model for prediction of the timing of stresses could be made both with and without dependence upon the preceding stresses. Any increase in predictive power in the case of stress-dependence would be evidence for the existence of stress-timing. A less sensitive statistical hypothesis involves auto-correlating the sequence of intervals between stresses. If there is dependence of the length of one such interval upon the immediately preceding one, then the autocorrelation of the sequence of intervals will show an abnormally high value when τ, the number of intervals the sequence is moved before being correlated with itself, is equal to one.

At the level of individual sentences a tendency toward equality of time intervals must produce changes in the pattern of time lapses in actual utterances. It must be found that otherwise structurally comparable segments of speech change their temporal character depending upon their rhythmic environment. One likely place to look for such changes would be sequences of major stresses on successive syllables, as in "big bug." If this phrase were imbedded in a sentence in which it is preceded by one unstressed syllable ("I sáw a bíg búg"), the time between the last two stresses should be less than if the embedding sentence has two preceding unstressed syllables ("Í saw a bíg búg"). This would be true at least if as Classe and Halliday suggest, the preceding interval is larger when there are two syllables than when there is only one.

1.3.3 Statement of Purpose

All of the above tests require that the time of occurrence of the stress be known. Several investigators (Miyake, Hollister, Classe, Newcomb) have shown that both speakers and listeners can treat spoken prose and poetry as rhythmic and that the feeling of the rhythmic beat of the syllable occurs at roughly the same position in the syllable under many conditions of experimentation. It is the purpose of this thesis to describe as fully as possible the measurability of this syllable beat in the perceptual aspect of its occurrence. It has been suggested that behavior is different towards an externally perceived rhythm than towards a self-regulated one. One cannot generalize therefore, from the listener's judgments of rhythm to the speaker's. One requirement for the study was repeated measurements on comparable syllables so that reliable measures might be obtained. It is statistically difficult, however, to obtain duplicate utterances in conversational speech, and since this investigator wished to explore the temporal constraints of this kind of normal speech, the perceptual domain was implied. One cannot say the same thing normally many times, but one can listen to a recorded utterance over and over. Two measures of syllable beat were used. One was a motor behavior well known by now, that of tapping the finger in time to the rhythm of the utterance, or, equivalently, to the beat of the syllable; the other was the auditory task of moving a click around in the syllable until it matched the syllable beat. These tasks have been used before, but the results have always been approximate. Previous investigators were interested in moving on to higher levels of structure before either the lower levels or the measuring tool itself were well enough understood. The following questions were asked, therefore, regarding these two behavioral measures and their relations to English speech rhythm: 1) Do different listeners tap their fingers at the same point in time for a given syllable? If the answer is "yes," then that point is fairly well established as the moment of the syllable beat, given absolute synchronization of the tap with the beat. If the answer is "no," however, not only must sources of variation be identified, but the range of variability between listeners must be established in order to make statistical statements about time lengths.

2) Do different listeners place the click at the same point in the syllable? Variation in this task has implications more directly related

to the perceptual system. A close match in click placement but not in taps would indicate that listeners hear the beat at the same time, but cannot hit it.

3) Do listeners place the click where they tap? If they do not, then the two tasks yield different results and cannot be measuring the same phenomenon.

4) Are tap and click placements equally variable, or is one a "sharper" indicator of a point in time? It is possible that one task is more appropriate for locating the rhythmic beat; listeners would respond to this appropriateness by giving more reliable responses in that task. The relative magnitudes of variation are useful in formulating hypotheses about the process of rhythm tracking, for if the variation of click placement were greater than the variation of tapping it would be less likely that the process of tapping involves a sequence of indentification of the beat followed by the actual tapping, with a corresponding addition of the errors of the two parts.

5) Are tap and click placements equally variable from syllable to syllable, or can the listener respond more reliably on some syllables? A lowering of variability for a given syllable might indicate that there is a more clearly defined beat there. This change in behavior should correlate with subjective feelings about the syllable.

The above questions all relate to the behavioral tasks as measures. A question which relates these measures to other levels of the problem of speech rhythm is:

6) With what physiologic-acoustic measures can the point of the syllable beat be correlated? Can anything more specific be said than has been said before?

CHAPTER II - Experiments

2.0 Introduction

The questions asked at the end of Chapter I involve the syllable beat in two ways. They ask first where it occurs in the speech wave or in absolute time, and second what is its nature as a behavioral stimulus. The second question is much more complex than the first and presupposes the first for a complete answer; in the light of previous studies on the syllable beat, however, even the simplest questions require better answers. Both Miyake (1902) and Classe (1939) used distributions of measures in determining the location of the syllable beat, but since they asked no "Yes-no" questions, the number of points in their distributions was not critical. They sought to locate the beat in the acoustic wave, and they imposed no *a priori* constraints on the accuracy of this location. This kind of open question of beat location is exemplified by Question 6 (p. 29) which asks for a clarification of the connection between syllable beat and acoustic wave. But Questions 1 through 3 ask questions about the exact location of the syllable beat for different listeners, and it is because of this possibility of separation that distribution size is important. Questions 4 and 5 ask similar exact questions about the variability of different subject's responses to different syllables. Question 6 does not immediately suggest a refutable statement, while Questions 1 through 5 have yes or no answers. Questions 1 through 5 can be translated into statistical terms by equating "location of the syllable beat" with "mean of the distribution of responses (taps or click placements) for that syllable" and "variability of response" with "variance of the distribution of responses." The questions then become statistical hypotheses about differences between means and variances of distributions. Much is known about the testing of such hypotheses, and the design of the experiments reported in this chapter is constrained not only by the desire to give adequate separation between different subjects' responses but also by the necessity of controlling for serial order effects (e.g. learning) and of completing the design so that conclusions are equally valid for all subjects and all syllables.

Two experiments were carried out, as suggested by the questions in Chapter I. In the first experiment, the behavioral task was to tap the finger to the beat of the syllable. Several finger taps to a single syllable generated a distribution of locations; the mean and variance of this distribution were used as measures to define the location of the syllable beat and the "rhythmicalness" of the syllable, respectively. In the second experiment, a movable click was placed in the syllable by the subject at a location where he felt he would have tapped, had he been tapping as in the first experiment. Again, a distribution of click locations was obtained for each syllable and again the mean and variance of this distribution were used as measures of beat location and rhythmicalness. The results of the two experiments were then compared to see if they were measuring like quantities. The numerical results of the experiments and the application of these results to Questions 1 through 5 of Chapter I are discussed in Chapter III. Applications of these results to the speech wave, rhythm, and the question of stress-timing in English comprise Chapter IV.

2.1 Experiment 1

2.1.1 Subjects and Speech Material

Preliminary studies indicated that subjects differed substantially in the location of their taps to a given syllable. Since the mean of each subject's taps to a syllable was to be compared with the means for other subjects on the same syllable, it was necessary to choose beforehand a number of subjects and a number of taps for each subject that would show the above-mentioned differences statistically. Since there was considerable variability between subjects, a small number of subjects was sufficient. Because a small number of subjects could be used, it was further possible to represent each subject's speech in the stimulus materials. In this way, each subject was able to react to his own as well as the other subjects' speech, and some control of idiolectal differences was obtained. Three subjects were engaged in conversation in a sound-proof recording room and the resulting speech was recorded through an Altec 633A boom-mounted microphone on an Ampex 350 tape recorder. From the resulting hour-long tape recording three utterances were chosen from each of the three subjects' speech according to the following constraints:

(1) Each utterance should be bounded at both ends by a major rhythmic juncture, i.e., a "long" period of silence or a linguistic pause. This rhythmic boundedness is desired so that the utterance is as rhythmically "complete" as possible, that is, the utterance contains most of the information about the location of the rhythmic beats. Such a criterion is necessarily judgmental, since little is known of the acoustic nature of rhythmic juncture.

(2) There should be some variation in rhythmic structure over the several utterances. The metric structure and the strictness of the

rhythm again had to be evaluated subjectively by the experimenter since the absolute measurement of these attributes was not possible at the time and was one of the major reasons for the experiment in the first place.

(3) There should be a broad sampling of phonetic types of syllable onset in the rhythmically accented syllables. The studies reviewed in Sections 4 and 5 of Chapter I indicated that the position of the beat is a function of the articulatory structure of the beginning of the syllable.

(4) Each utterance should be fairly even in loudness, so that audibility is not affected in shifting attention from one part of the utterance to another. This criterion was not always easy to meet, as can be seen from the speech power record of utterance #7 (see Appendix A), where the last two words are spoken softly (this tendency toward lowering the voice at the end of a phrase was characteristic of this subject's speech). Overall speech level was normalized when tape loops were made for experimental purposes.

(5) Utterances should be long enough that interesting rhythmic patterns are derivable from them but they should not be so long that it would be inefficient to play the entire utterance for a single response.

(6) There should be a fairly even distribution of the various characteristics of interest (e.g., rhythmic structure, phonetic onset types) over the three idiolects, making as complete a statistical design as possible.

Spectrograms and mingograms showing the speech signal and speech power of the nine utterances are given in Appendix A. Each utterance was recorded on a loop of 1/4 inch magnetic tape (Scotch Instrumentation #188) approximately 30 inches in length. Exact tape loop lengths in sec for a single revolution at 7 1/2 inches per sec (i.p.s.) are given in Table 2.

TABLE 2

Loop Lengths in Milliseconds (at 7 1/2 i.p.s.)

Before and After Presentation in Experiment 1

		Subject 1	Subject 2	Subject 3
Loop 1	before	4002	4003	4002
	after	4001	4002	4002
Loop 2	before	4003	4002	4002
	after	4005	4003	4003
Loop 3	before	4006	4006	4007
	after	4006	4007	4007
Loop 4	before	4004	4003	4004
	after	4004	4003	4005
Loop 5	before	4002	4002	4001
	after	4002	4000	4002
Loop 6	before	3996	3999	4001
	after	3998	3999	4001
Loop 7	before	4003	4000	4000
	after	4002	4000	4000
Loop 8	before	4002	4004	4004
	after	4003	4004	4004
Loop 9	before	3999	4001	3999
	after	4001	4001	3998

2.1.2 Experimental Apparatus

Figure 1 shows in block diagram form the experimental apparatus. Each utterance was recorded on one track of a two-track loop of magnetic instrumentation tape approximately 30 inches long. The actual lengths are given in Table 2.

On the other track of each loop, positioned a few inches before the onset of the utterance, was a click produced by touching the tape with a magnetized razor blade. Subjects heard the speech from track 2, but not the click from track 1. The form of each click was a single oscillation whose positive and negative peaks were greater than 1 volt d.c. in amplitude and approximately 1 msec apart. This pulse was used to reset the frequency timer (hereafter referred to as the "clock") by initiating a fixed length pulse in a pulse generator. The onset of the generated impulse stopped the clock which was then read by a printer. The subsequent offset of the pulse restarted the clock. For this reason, this pulse will be called read-time pulse. This scheme was used to stop and start the clock because the time taken by the printer to read and restart the clock was unknown and not easily measurable. The length of the read-time pulse could be calibrated easily, however, by reversing the polarity of the start and stop switches on the clock, starting at the pulse onset and stopping at the offset. On the finger with which subjects tapped they wore a copper thimble which was connected, in series with a 20 kilocycle[2] oscillator and a copper plate, across the input terminals of the impulse generator. The copper plate was fixed to the table part of the desk-chair in which the subjects sat, and when they tapped on the copper plate, the 20 KC signal initiated an impulse

[2]This value was chosen to avoid a beat tone, audible through the earphones, that occurred near 10 KC and 100 KC.

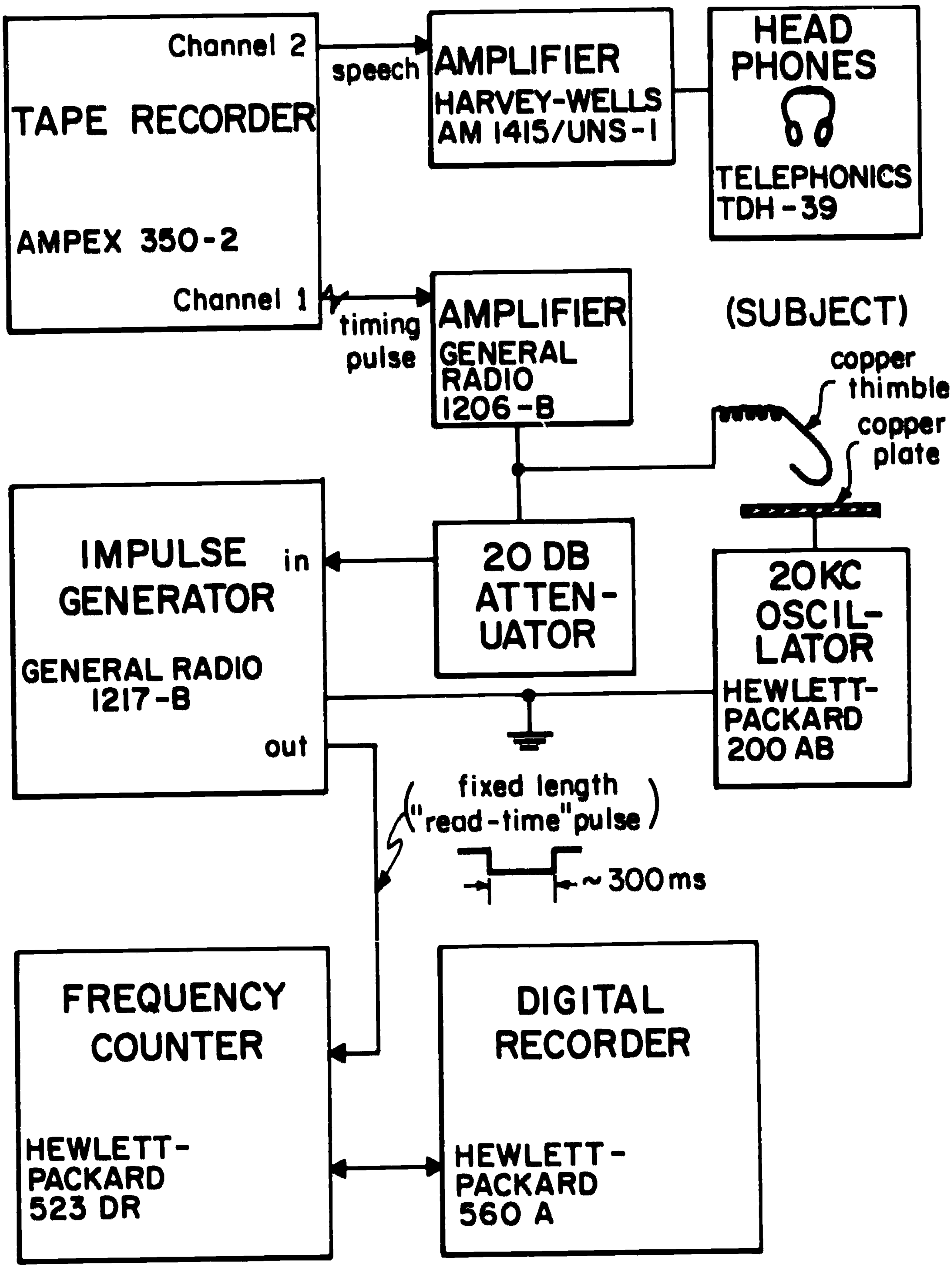

Fig. 1. Apparatus for Experiment 1

in the impulse generator. The time for a single revolution of the tape loop, during which the subject tapped once, can therefore be described as the sum of four times: 1) the length of the read-time pulse initiated by the timing pulse on track 1 of the tape loop; 2) the time between the offset of the read-time pulse and the time of the subject's tap; 3) the length of the read-time pulse initiated by the tap; and 4) the time between the offset of this second read-time pulse and the onset of the read-time pulse initiated by the timing-pulse for the next revolution. As far as this investigation is concerned, the time period of interest is the sum of the first two times, i.e., the time between the occurrence of the timing pulse and the subject's tap.

2.1.3 Experimental Design

Subjects were seated in a desk-chair in a sound-proof room and were asked to tap their finger in time to the beat of a given orthographic syllable. (Exact instructions are given in Appendix D.) "Syllable," "syllable beat" and the nature of the task were understood intuitively and immediately by all subjects. Once they started tapping to a given syllable, they continued to tap to that same syllable until the utterance stopped playing. During each revolution of the tape loop, therefore, the subject heard the utterance once and tapped once to the syllable in question. The experimenter counted the number of revolutions and controlled the number of taps given by the subject for the syllable by turning off the sound. Subjects were given a printed version of the utterances with all but the first few syllables numbered in some order. The early syllables were not marked for tapping because a tap would be affected by the subject's reaction time and therefore would not mark the syllable beat. Thus, if there were ten syllables in the printed version of the utterance and subjects were to tap to the last eight of these, the numbers one through eight were written over the vowels of the last eight orthographic syllables in a sequence such that various order effects were controlled. The subject listened to the first playing of the utterance and then tapped once to the syllable marked "1" on each succeeding playing until the utterance was turned off. When the utterance was turned on again, the subject listened to the first playing and then tapped to the syllable marked "2" until the utterance again was turned off. Each experimental session consisted of a subject tapping to all marked syllables of a given utterance. The marked syllables, the order of syllable numbering, the number of taps to each syllable and the order of presentation of the nine utterances for the three subjects are given in Table 3. It will be seen from Table 3 that each subject tapped to every marked syllable, and that the number of taps by the three subjects to a given syllable was the same, but that the number of taps to different syllables was different; this was done to see whether overall tapping behavior depended upon the number of taps. Further, the order of the syllables was different for different subjects. Within the ordering of the syllables for tapping there was an attempt to alternate "rhythmic" and "non-rhythmic" syllables. A syllable was categorized as "rhythmic" if the experimenter felt that it marked a strong rhythmic beat in the utterance and thus was probably easier to locate and tap to than the other "non-rhythmic" syllables which were passed over more quickly in the speech. It was decided that alternation of hard and easy tasks, if indeed they turned out to be so, would keep the subject more interested in the experiment and his tapping behavior more consistent.

The number of taps per syllable per subject was chosen as 50 or 100 with 50 as the more common. Preliminary data had indicated that a

TABLE 3

Syllable Presentation Schedule for Experiment 1

Subject 1 (Number of Taps per Syllable

Shown Below the Syllable)

```
Order of Presentation   Order of Presentation of Syllables
  of Utterances
                    9     6    4    7    1    2    5        8    10      3
   1     Use the weight of the line to get more_1 and more_2 out
                    50    50   50   50   50   50   50       50   100     50

                    6     3    4    9    2    7    5     1     8
   6     Spinners are   par tic  u   ly  good in  cur rents
                    50    100  50   50 100   50   50   100    50

                   6     1       3     8     5     2  7  4      9
   8     See when he reared back and fired it by  a    guy
                   50    50      50   100    50    50 50 100    50

                2    5     4    3   1
   9     After the play was  o  ver
               100  100  100  100  100

             1      6    3   8     2    5    4   7
   2     He wants to  be  a    per for mer now
             50     100 50 100    50  100   50  50

                      3    10   5    8    4     9    2    6    1     7
   4     Talks in terms of get ting a  name for him self  so
                      50   50   50   50   50    50   50   50   50   100

                     5    7  1     6    9     4     8        3      2
   5     I would like to know how they con struct these things
                     50   50  50    50  100   50    100      50     50

              9    2      7      6      4    1     10   8     5       3
   7     I_1 like pre_1 dict_1 ing  I_2 love to pre_2 dict_2 things
              50   50     50     50   100   50    50   50    50      50

                      1    8    3    6      9    2   5   11   7  4    10
   3     What will hap pen now and sort of go out on a   limb
                      50   50   50   50    100  50  50   50  50  50   50
```

Subject 2 (Number of Taps as for Subject 1)

```
Utterance order            Syllable order
                       1      6   8     3  10   9    5      4    2      7
   5     Use the weight of the line to get more_1 and more_2 out
                         9    6    7  3  5  1     8   4    2
   7     Spin nere are par tic u ly good in cur rents
                        4    3     5     2     9    8  1  6  7
   3     See when he reared back and fired it by a guy
                    4     1    2   3   5
   1     After the play was o ver
                 3     2  7 6   4    1    8    5
   6     He wants to be a per for mer now
                        9    2    7     4 10    1    6    8    5    3
   8     Talks in terms of get ting a name for him self so
                        8    7  9    6    4    2     3      1     5
   9     I would like to know how they con struct these things
                  1      6     3     8  10   5    2    4     7     9
   2     I_1 like pre_1 dict_1 ing I_2 love to pre_2 dict_2 things
                            4  2   6   11   3   5  10   9   1 7  8
   4     What will hap pen now and sort of go out on a limb
```

Subject 3 (Number of Taps as for Subject 1)

```
Utterance order            Syllable order
                        9      8  4   1    6    5    7      2    10     3
   9     Use the weight of the line to get more_1 and more_2 out
                          3    6    1  9  5   7    2    4    8
   2     Spin nere are par tic u ly good in cur rents
                        2    9     7     6     1    4  5  8  3
   4     See when he reared back and fired it by a guy
                     2    3    4  1  5
   5     Af ter the play was o ver
                 7    4  5   2  8   3   6    1
   7     He wants to be a per for mer now
                          1    6  3   10  2    5    8    4   7   9
   3     Talks in terms of get ting a name for him self so
                         1   4  2    3    6    8    5      7     9
   1     I I would like to know how they con struct these things
                 3   10     5    2  8    9  4   6      1      7
   6     I_1 like pre_1 dict_1 ing I_2 love to pre_2 dict_2 things
                           10   6   8   2   7   11  1  4  5  9  3
   8     What will hap pen now and sort of go out on a limb
```

subject's 50 taps to a syllable would give a distribution whose mean was statistically stable enough to give adequate separation from the means of the distributions of 50 taps by the other subjects on the same syllable. The selection of syllables to be tapped to 100 times was based on two criteria, first that there be an even overall sampling of the different rhythmic types of syllable, from the most rhythmically stressed to the least, and second that there be a consistent time length for the experimental sessions. The first criterion specified which combinations of syllables might be tapped to 100 times; the second criterion forced greater numbers of these syllables to be chosen from utterances with fewer syllables. The total number of taps to an <u>utterance</u> thus was kept relatively constant. If the numbers of taps to the syllables of an utterance are added, the result is 550 taps for all utterances except the fourth, which has a total of 500, and the ninth, with a total of 600. Each experimental session lasted approximately 45 minutes, with a rest break about half way through, upon the request of the subject.

Each subject tapped to all the marked syllables of a single given utterance in one experimental session, and so there was probably little interaction among utterances. Nevertheless the presentation schedule of the nine utterances was different for the three subjects with no subject hearing any two utterances one after the other in the same order, and no two utterances from the same dialect in successive sessions.[3]

[3]That is, if a subject heard utterance A in one session and utterance B in the next session, this implies that utterances A and B were from different idiolects and that no other subject heard B in the session immediately following the one in which he heard A.

2.2 Experiment 2

2.2.1 Speech Material

Preliminary experimentation indicated that the task of matching a click with a syllable beat is a time consuming one and so a subset of the speech material for Experiment 1 was used in Experiment 2. One utterance was chosen from each of the three idiolects with an attempt to obtain different rhythmic and syllable onset types as before. The utterances chosen were numbers 2, 5, and 7.

For two reasons, a separate tape loop of the entire utterance was made for each syllable of that utterance. In Experiment 1 each tape loop was played approximately 2000 times, and some decrease in the signal-to-noise ratio was noted. It was felt that this signal degradation might prove important in the more strictly auditory situation of Experiment 2. The total number of revolutions for a single syllable-loop in the second experiment ranged between 221 and 644 with an average of 554; no signal degradation was noted.

The second reason for separate tape loops relates to the time interval limits of the variable click described in the next section. In order to give subjects a satisfactory degree of control over click placement, the total click range was set at approximately 1.5 sec. With a greater range of placement, subjects were not able to control local accuracy adequately; with a smaller range, a relatively large turn of the knob produced no perceivable change in location. Since the utterances were greater than 1.5 sec in length, not all of the syllables in an utterance would fall in this 1.5 sec range; the click could not be placed on the syllables outside the range. Thirty-one tape loops were made and a timing pulse was placed on each loop with a magnetized razor blade as in Experiment 1, but so that the syllable with which the loop was associated fell well within the limits of the 1.5 sec range. Subjects could thus move the click so that it both clearly preceded and clearly followed the syllable in question.

2.2.2 Experimental Apparatus

Figure 2 shows in block diagram form the apparatus for Experiment 2. Subjects heard the utterance from track 2 of the tape loop. The timing pulse on track 1 simultaneously started the clock (frequency timer) and initiated a variable length pulse in the circuit labelled "Time-Variable Click". (The componentry of this box is given in Appendix B.) The offset of the variable length pulse produced a sharp click which was fed through an amplifier to both the earphones and the stop input of the clock. The length of the variable pulse was adjusted by turning the knob on the box. The subject could thereby move the audible output click around in the syllable until he was satisfied with its location. The interval between the timing pulse on track 1 and the audible click was recorded by the printer at each revolution of the tape loop. In this way the successive locations of the click were measured relative to the timing pulse.

2.2.3 Experimental Design

The subjects for Experiment 2 were two of the three used in Experiment 1, the third being unavailable. Subjects sat at a table in the sound-proof room and were asked to move the click until they judged that it coincided with the time when they would have tapped, had they been tapping to a given syllable. (Exact instructions are given in Appendix D.) Since this task was more auditory in nature there was not

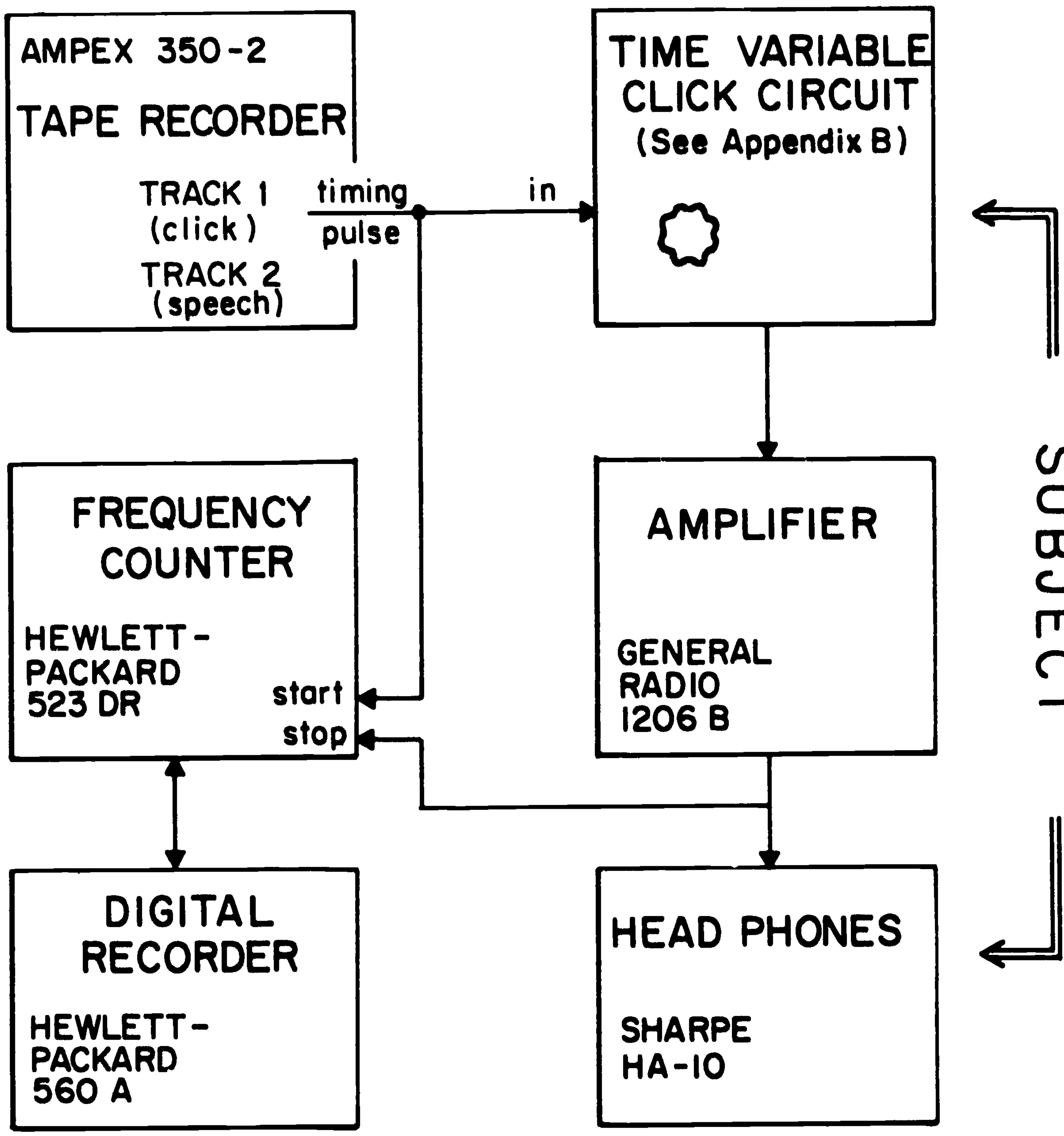

Fig. 2. Apparatus for Experiment 2

necessarily a reaction time component to the click placement, and so all syllables of the utterance were marked for click placements. The principle reason for wanting measurements on the earlier syllables was that some of these earlier syllables were rhythmically stressed and established a beginning time reference point for the sentence rhythm. No measurement of them had been possible in the tapping experiment. Subjects were given a sheet of paper on which were printed the successive utterances they were to hear. At each trial (that is, each satisfactory placement of the click) the subject reacted to a syllable in an utterance different from the one which he heard on the trial before. Syllables were again presented in a rhythmically balanced order to help the subject retain interest in the task. He heard the utterance as many times as he wished, moving the click between hearings. Subjects were instructed to move the click to both a very early and a very late position during the first few revolutions of the loop so that they were sure that the click could both precede and succeed the preferred location. This instruction was given to minimize the probability that the subject would move the click in towards the location from one side or the other, never quite reaching the preferred spot. This manner of response would give measurements biased toward the last direction from which the click was moved. It remains quite probable that some of the obtained measurements have this kind of bias, for example, in a sequence of placements where the final location is either earlier or later in the speech than any of the other locations on the trial. More will be said of this problem in Chapter IV.

Three different conditions of stimulus presentation were employed in an attempt to control for perceptual asymmetries of the two ears. Ladefoged and Broadbent (1960) and Fodor and Bever (1965) noticed statistical biases in click placement between the two ears in somewhat similar tasks. The three conditions were, therefore, both click and speech in both ears, click in left ear and speech in right, and click in right ear and speech in left. It will be noted that there may be a fundamental difference in the perceptual matching process between the first condition, where click and speech form a single signal and the last two conditions, where the click and speech come in different ears and must be matched at a higher order auditory center. The order of stimulus conditions was also balanced to control for order effects.

The schedules of utterance, syllable and auditory condition presentation are given in Table 4. Five click placements were obtained for each subject on each syllable under each of the three conditions, for a total of fifteen placements per subject per syllable.

statistical biases in click placement between the two ears in somewhat similar tasks. The three conditions were, therefore, both click and speech in both ears, click in left ear and speech in right, and click in right ear and speech in left. It will be noted that there may be a fundamental difference in the perceptual matching process between the first condition, where click and speech form a single signal and the last two conditions, where the click and speech come in different ears and must be matched at a higher order auditory center. The order of stimulus conditions was also balanced to control for order effects.

The schedules of utterance, syllable and auditory condition presentation are given in Table 4. Five click placements were obtained for each subject on each syllable under each of the three conditions, for a total of fifteen placements per subject per syllable.

TABLE 4

Click Experiment Stimulus Syllable Schedule

	SCHEDULE I						SCHEDULE II					
	Session 1		Session 2		Session 3		Session 1		Session 2		Session 3	
Trial	Syl.	Cond.	Syl.	Cond.	Syl.	Cond.	Syl.	Cond.	Syl.	Cond.	Syl.	Cond.
1	8	1	4	1	16	1	3	1	25	1	21	1
2	19	2	21	2	21	3	30	2	1	2	1	3
3	25	3	8	3	20	1	20	3	30	3	18	3
4	15	1	25	1	4	2	24	1	3	2	25	2
5	31	2	16	2	12	3	16	2	12	3	14	2
6	6	3	6	2	8	2	5	3	5	2	30	1
7	22	1	20	3	25	2	12	1	24	3	3	3
8	11	2	31	1	6	1	9	2	18	1	24	2
9	16	3	12	2	31	3	29	3	29	2	5	1
10	30	1	22	3	17	1	2	1	14	3	15	1
11	20	2	11	1	11	3	18	2	9	1	9	3
12	1	3	30	2	13	1	31	3	15	2	29	1
13	12	1	1	1	22	2	14	1	2	3	19	1
14	3	2	17	2	14	1	22	2	31	1	31	2
15	28	3	28	1	30	3	10	3	10	2	13	3
16	9	1	13	3	1	2	26	2	19	3	2	2
17	23	2	3	1	28	2	7	2	22	1	17	2
18	17	3	14	2	3	3	15	3	13	2	10	1
19	5	1	9	3	18	3	4	1	26	3	22	3
20	13	2	23	1	9	2	28	3	7	1	7	3
21	26	2	5	2	23	3	19	2	28	2	26	1
22	14	3	26	3	5	3	23	1	4	3	4	2
23	24	1	18	1	26	1	13	1	17	1	28	1
24	2	2	2	1	2	3	8	2	8	1	8	3
25	29	3	19	3	24	2	17	3	23	3	23	2
26	7	1	24	3	7	2	11	1	20	2	11	2
27	18	2	15	2	29	1	27	2	27	3	27	1
28	27	3	29	2	10	1	6	3	16	1	6	2
29	9	1	7	3	27	2	20	1	11	3	21	3
30	10	2	27	1	4	3	21	2	12	2	1	1
31	15	3	10	3	21	1	16	3	6	1	25	3

Scheduling

Subject 1: II, I, II, I, II

Subject 2: I, II, I, I, II

CHAPTER III - Experimental Results

3.1 Introduction: Reliability and Validity

The two attributes of a statistical tool of most immediate interest are its reliability and its validity. A measure is reliable if it gives consistent answers to the same question. It is valid if its answers bear some desired relationship to the question. The present chapter will concern itself in the main with the reliability of the two behavioral tasks used in the two experiments described in Chapter II. Chapter IV will treat the validity of these two tasks as measures of syllable beat and speech rhythm. Since the present chapter contains statistical material, many readers will not be familiar with some of the notions used. Appendix C is an expanded explanation and motivation of these notions and their use. The numbering of the sub-sections of Appendix C corresponds to the sub-sections of this chapter. Thus, section C.3.1 amplifies the discussion of reliability and validity treated here in section 3.1.

3.2 Some Statistical Considerations

3.2.1 Mean and Variance

If a subject's responses form a distribution of points drawn from a potentially continuous domain, such as time in this case, a measure of central tendency is desired as a single datum to represent the entire distribution. For various reasons the expected value, or mean (symbolized "m"), of the distribution was chosen as this measure of central tendency. Likewise, as a measure of dispersion or variability of responses, the second central moment, or variance (symbolized "s^2"), was chosen. In the language of the questions at the end of Chapter I, if a subject taps his finger fifty times to the beat of a given syllable, each tap generating a time interval which is recorded as a datum, then the mean of these time intervals is taken to be the "place where the subject taps" and the variance of the fifty intervals is the "variability with which he taps."

3.2.2 Sequential Dependency in the Data and Its Effect on Variance

From the point of view of probabilistic variation it is important that the distributions of taps be considered random samples from some underlying population or populations. But it is clear that the characteristics of samples may not be independent of each other, and even within a single sample, a subject's tap may depend not only on the speech he is listening to, but also on other taps he has made. For example, he might decide he had tapped too soon last time and adjust his movements to compensate; or, he might learn better and better where he ought to tap and his later taps might thereby be more tightly clustered than his earlier taps. Since both the location and dispersion of the subject's responses to a syllable are to be used as measures bearing on the rhythm of the utterance, it is important to decide whether the location and dispersion change to any great degree over the time that the distribution is gathered. One way in which any sequential dependency of the distribution of taps can be discovered is auto-correlation of the sequence. The discrete auto-correlation function, $\phi(\tau)$, for values of τ from 1 to 30 averaged over the 243 subject-by-syllable distributions, is given in Table 5.

Table 5

Discrete Autocorrelation Function (ϕ) of

Tapping Responses

τ	$\phi(\tau)$	τ	$\phi(\tau)$
1	.17	16	-.01
2	.09	17	-.02
3	.07	18	-.03
4	.05	19	-.02
5	.04	20	-.02
6	.03	21	-.01
7	.02	22	-.02
8	.01	23	-.01
9	.01	24	-.01
10	.00	25	-.00
11	-.00	26	.02
12	-.02	27	.00
13	-.02	28	.00
14	-.01	29	.00
15	-.02	30	.00

The values of ϕ (τ) for low values of τ show that there is a slight dependency of successive taps, and that this dependency exists over as many as ten responses. Since $\phi(1)$ is positive, subjects tend to tap close to where they tapped the time before, giving fairly long oscillations of behavior around some central point. The meaning of the distribution mean as a measure of central tendency is made less clear by these long-range shifts in tapping behavior. These shifts in tapping make the distribution mean less stable then if there were no shifts. Means were computed for the first and second halves of the tapping sequences. The variance of these means was significantly greater than would have been expected using the entire distribution variance, s^2, as a comparison. Variances were also computed for the first and second halves of the tapping distributions. Subjects' taps show a variability decreasing over time, as indicated by the number of syllables on which the second-half variance was lower than the first-half variance. Table 6 gives these frequencies for three subjects.

Long range trends inflate variance, so a measure of variability that is insensitive to long-range trends was computed for the distributions. This measure, the so-called "mean square successive difference", symbolized d^2, is the average of the squared difference between successive data points. The relation between s^2 (variance) and d^2 is summarized in Table 7. It shows d^2 to be very highly correlated with s^2 in relative size, from one syllable to another.

Table 6

First-Half versus Second-Half Variance Size Comparison

for Tapping Distributions, Types A & B Syllables

	1st Half Variance Greater	2nd Half Variance Greater	Total
Listener 1	25	13	38
Listener 2	26	12	38
Listener 3	25	13	38
Total	76	48	114

$$\chi^2_{2d.f.} = 12.7, \ p < .005$$

Table 7

Correlation of s^2 with d^2 and Log (s^2) with Log (d^2)

Over the Syllable Tapping Distributions

	Subject 1		Subject 2		Subject 3	
	r_{s^2, d^2}	$r_{\log s^2, \log d^2}$	r_{s^2, d^2}	$r_{\log s^2, \log d^2}$	r_{s^2, d^2}	$r_{\log s^2, \log d^2}$
All Syllables	.88	.94	.95	.92	.93	.90
Utterance 1	.96	.97	.98	.93	.80	.78
Utterance 2	.85	.94	.88	.78	.98	.98
Utterance 3	1.00	.98	.94	.93	.93	.92
Utterance 4	.96	.96	.98	.95	.99	.97
Utterance 5	.95	.83	.99	.99	.96	.91
Utterance 6	.89	.91	.76	.85	.98	.97
Utterance 7	.96	.97	.94	.93	.87	.81
Utterance 8	.91	.92	.97	.94	.92	.95
Utterance 9	.90	.94	.98	.98	.95	.94

$$r_{xy} = \frac{\sum_{i=1}^{N} (x_i - \bar{X})(y_i - \bar{Y})}{N \times S_x \times S_y},$$

where $\bar{X}$ is the mean of the x_i

S_x is the standard deviation of the x_i

and N is the number of syllables in the utterance.

3.2.3 Form of the Underlying Distribution

It is an important underlying assumption of many widely used statistical tests that the data being analyzed be a sample drawn from a normally distributed parent population. The forms of the distributions of tap and click placements were tested for normality. The results of the comparison of the tapping data with a normal distribution are given in numerical and graphic form in Table 8 and Figure 3. A chi-square test of goodness-of-fit on the twenty-four .25 sec intervals shows the distribution to be significantly non-normal, but the fit is close enough to justify making probability statements based on its being normal. Such probabilities might perhaps be in error in the second or third significant digit.

Because of the very close fit of the distributions to normality using s^2 as a measure of dispersion, and because of the high correlation of s^2 with d^2, s^2 was chosen as the measure of dispersion for the purposes of this chapter.

The distributions of click-placing data were tested for normality in two ways. Within each subject's responses to a given syllable there were three conditions of stimulus presentation. For subject 1 there was a highly significant difference in click placement for the three conditions, with the placements for condition 3 (speech in left ear, click in right) earlier than those for condition 2 (speech right, click left), and with condition 1 giving intermediate results. Subject 2 showed no such difference among the three conditions. The click data were therefore grouped in two ways. In the first grouping, the five responses for each condition on each syllable for each subject (subjects-by-syllables-by-conditions) were taken as a single sample. The second grouping ignored differences according to the three conditions and grouped all

Table 8

Form of the Tapping Distributions Compared with Form of the Standard Normal Distribution - N = 9720

u	Cumulative Distribution Functions, F(u) Normal	Data	Density Functions, $f_{u_i, u_{i-1}} = F(u_i) - F(u_{i-1})$ Normal	Data
-3.00	.001	.001	.001	.001
-2.75	.003	.004	.002	.002
-2.50	.006	.009	.003	.005
-2.25	.012	.016	.006	.007
-2.00	.023	.026	.011	.010
-1.75	.040	.043	.017	.017
-1.50	.067	.068	.027	.026
-1.25	.106	.103	.039	.035
-1.00	.159	.153	.053	.049
- .75	.227	.219	.068	.067
- .50	.309	.307	.082	.088
- .25	.401	.406	.092	.099
.00	.500	.513	.099	.106
.25	.599	.612	.099	.099
.50	.691	.709	.092	.097
.75	.773	.784	.082	.075
1.00	.841	.850	.068	.066
1.25	.894	.902	.053	.052
1.50	.933	.934	.039	.032
1.75	.960	.958	.027	.024
2.00	.977	.974	.017	.016
2.25	.988	.985	.011	.011
2.50	.994	.991	.006	.006
2.75	.997	.995	.003	.005
3.00	.999	.998	.002	.003
$+\infty$	1.000	1.000	.001	.002

χ^2 Goodness of Fit Test on Density Functions Yields

$\chi^2_{25df} = 89.8$
($p < .001$)

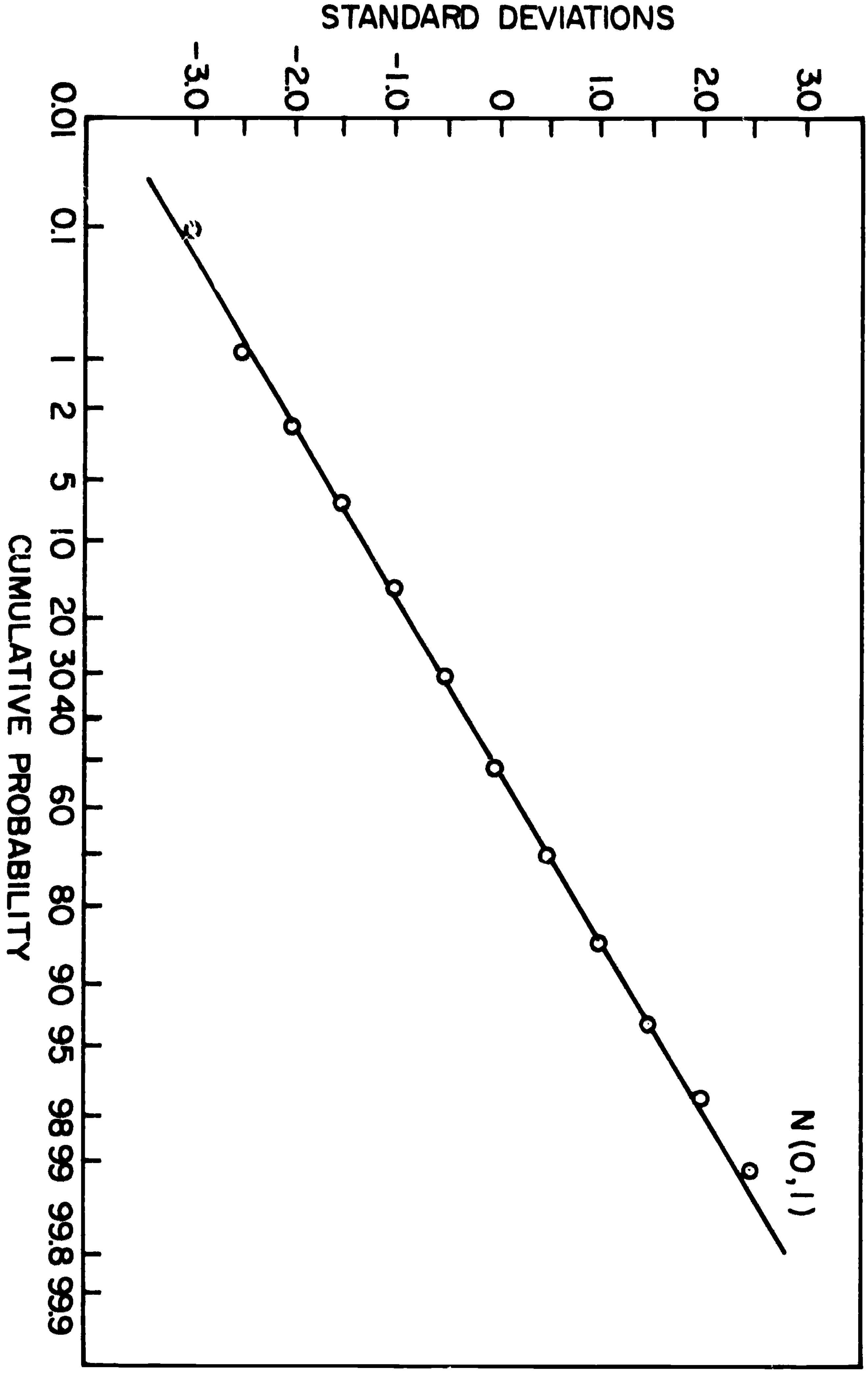

Fig. 3. Comparison of Cumulative Distribution Functions of Tapping Data (o) and Standard Normal (-)

fifteen responses of a subject to a given syllable. The pooled cumulative distribution functions were then compared with Student's t-distribution with four and 14 degrees of freedom, respectively. The results for four degrees of freedom are given in Table 9 and Figure 4. In neither case does the obtained distribution match the predicted t-distribution. Since the effect of the three conditions was different for the two subjects, the "by five" and "by fifteen" groupings were carried out for the two subjects separately. The two pairs of distributions were then compared, with the result that the "by five" distributions were quite close (χ^2 goodness-of-fit test with 15 degrees of freedom, $\chi^2_{15} = 10.4$), but the "by fifteens" distributions were very different ($\chi^2_{21} = 42.6$). The density functions and χ^2 calculations are shown in Table 10. This inequality in matching of the "by fifteens" distributions may be a function of the unequal differences in behavior on the different conditions by the two subjects. An hypothesis about the underlying cause of the extreme rectangularity of this distribution will be discussed in Section 2 of Chapter IV.

3.2.4 Error from Experimental Apparatus

The experimental apparatus contributed some error to the measurements, and care was taken to make sure the error variance was well below the subjects' variance in order of magnitude. The lowest variance on any syllable by any subject in the tapping experiment was 3 $csec^2$. The order of magnitude of the error variance of loop rotation time was 1.5 $msec^2$. Other error components were unmeasurably small. It was assumed that time changes owing to loop rotations were uncorrelated with differences in the subjects' behavior.

Table 9

Form of the Click Distributions Compared with Form of

The Student's t-Distribution with 4 Degrees of Freedom

	Cumulative Distribution Functions, $F(u)$ Student-t	Data	Density Functions, $f_{u_i,u_{i-1}} = F(u_i) - F(u_{i-1})$ Student-t	Data
-2.00	.058	.001	.058	.001
-1.75	.078	.039	.019	.038
-1.50	.104	.092	.026	.053
-1.25	.140	.148	.036	.056
-1.00	.187	.233	.047	.086
- .75	.248	.324	.060	.091
- .50	.322	.416	.074	.091
- .25	.408	.519	.086	.103
.00	.500	.601	.092	.082
.25	.592	.683	.092	.082
.50	.678	.767	.086	.083
.75	.752	.833	.074	.067
1.00	.813	.896	.060	.062
1.25	.860	.947	.047	.051
1.50	.896	.997	.036	.050
1.75	.922	1.000	.026	.003
+ ∞	1.000	1.000	.078	.000

N = 900

χ^2 Goodness of Fit Test on the Density Functions Yields

$\chi^2_{16df} = 138$

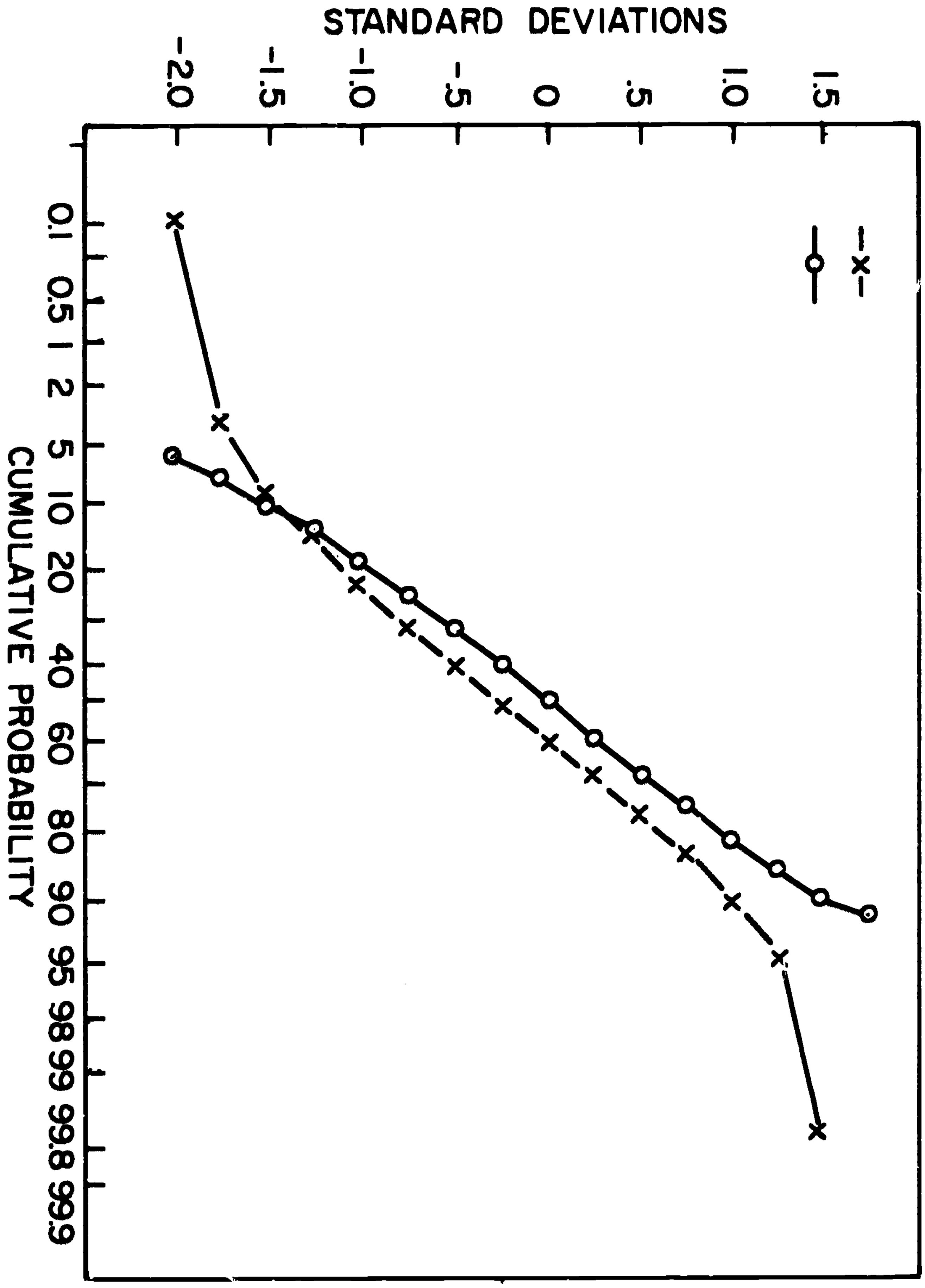

Fig. 4. Comparison of Click Distributions (-x-) with Student's t-Distribution with 4 Degrees of Freedom (-o-)

Table 10

Comparison of "By Five" and "By Fifteen" Distributions

	"By Fives" Distributions		"By Fifteens" Distributions		
	Subject 1	Subject 2	Subject 1	Subject 2	
-2.75			.000	.002	
-2.50			.007	.000	
-2.25			.000	.004	
-2.00			.009	.011	
-1.75	.000	.002	.007	.004	
-1.50	.038	.038	.027	.049	
-1.25	.049	.058	.056	.016	
-1.00	.049	.062	.058	.082	
- .75	.093	.078	.067	.073	
- .50	.100	.082	.080	.071	
- .25	.102	.080	.102	.107	N = 450
.00	.102	.104	.096	.082	
.25	.084	.080	.104	.102	
.50	.080	.084	.100	.058	
.75	.071	.096	.069	.089	
1.00	.062	.071	.056	.084	
1.25	.056	.069	.049	.060	
1.50	.053	.049	.051	.049	
1.75	.053	.047	.020	.022	
2.00	.007	.000	.022	.018	
2.25			.007	.013	
2.50			.016	.002	

$X^2_{15df} = 10.35$ ($p > .5$) $X^2_{21df} = 42.6$ ($p \approx .005$)

In the click placing experiment also, the entire measurable error contribution was from the tape loop; as in the tapping experiment, this was 1.5 $msec^2$. The smallest variance in click placement, for both subjects on all syllables, was 45 $msec^2$. The ratio of thirty between click placing variance and apparatus error is not as great as that for the tapping data but can affect variance ratios at worst in the third significant digit.

3.3 Questions about Differences Between Variances

Because of certain statistical considerations, it is necessary to answer questions about systematic variation of variance before means can be compared. Analyses of variance were carried out on the variances of the tapping and click distributions. These analyses were justified by the reasonable homogeneity of the variances within the two sets of distributions. In analyzing the tapping data, three factors were chosen, namely subjects (symbolized L, for "listener"), dialects (D), and syllable types (T). There were three levels of each factor. The three levels of L were the three subjects, and the levels of D were the utterances spoken by each of the three subjects. (The first three utterances belonged to the first level of D, the next three to the second and the last three to the third.) It had already been noticed that there was considerable correlation between subjects in the size of variance, depending on the syllable, with smaller variances associated with the more stressed or rhythmic syllables. All ninety-seven syllables were therefore "typed" a priori according to the role that the experimenter thought they played in the utterance in which they occurred. This typology will be treated further in Chapter IV. For the present analysis, it suffices to say that there are three general classes of syllable: syllables carrying major stress (type A), syllables carrying minor stress and/or rhythmic beat (type B), and non-stressed, non-rhythmic syllables (type C). The results of the analysis of variance on the logarithms of the variances of the tapping distributions and the average values for the main effects are given in Table 11. They show that there are significant differences in the size of variance 1) between the subjects (subject 2 < subject 1 < subject 3); 2) between syllable types (type A < type B < type C); and 3) interaction L x T. No significant variation was found to result from differences in dialect. These findings indicate that some subjects are more consistent tappers than others and that the consistency of tapping to a given syllable is related to the role of that syllable in the rhythm of the utterance.

A similar analysis of variance was carried out on the variances of the click placing distributions. Again three factors were considered, but in this analysis they were subjects (L), syllables (S) and conditions of stimulus presentation (C). There were two subjects and hence two levels of factor L. No systematic variation was immediately noticeable among the variances according to syllable type, and so each syllable was a separate level, making a total of 30 levels of factor S.[5] The three levels of factor C corresponded to the three conditions of stimulus presentation, namely both click and speech in both ears, click in left ear, speech in right, and click in right ear, speech in left. There were no findings of statistical significance from this analysis of variance, as reported in Table 12. An equivalent statement is that among the variances of the click placing distributions

[5] Syllable #18 was omitted from the analysis.

Table 11

Tapping Log-Variance Analysis of Variance Table

Source	Degr. Freedom	Sum of Squares	Mean Squares	F-Ratio	Probability Level
Listeners (L)	2	4.529	2.264	44.5	< .0005
Dialects (D)	2	.1811	.0905	1.78	.1 - .25
Syl. Types (T)	2	4.602	2.301	9.16	<.05
L x D	4	.1346	.0336	.66	>.5
L x T	4	.9443	.2361	3.58	∿.05
D x T	4	.3244	.0811	—	—
L x D x T	8	.5271	.0659	1.29	∿.25
Residual	216	10.99	.0509		

<u>Marginals</u> ($msec^2$)

Listeners -	L_1		1370
	L_2		1060
	L_3		2320
Dialects -	D_1		1630
	D_2		1500
	D_3		1370
Syllable Types -	T_1	(A)	1050
	T_2	(B)	1480
	T_3	(C)	2160

Table 12

Click Log-Variance Analysis of Variance Table

Source	Degr. Freedom	Sum of Squares	Mean Squares	F-Ratio	Probability Level
Subjects (L)	1	.0450	.0450	.27	>.5
Syllables (S)	29	5.075	.1750	1.04	>.25
Conditions (C)	2	.1040	.0520	8.26	N.S.
L x S	29	4.091	.1411	.84	>.5
L x C	2	.1187	.0593	.35	>.5
S x C	58	6.705	.1156	.69	>.5
Residual	58	9.781	.1686		

Marginals for Syllable Type ($msec^2$)

Types A & B - 857

Type C - 1119

there are no systematic variations resulting from differences between the subjects, syllables, conditions of stimulus presentation, or any two-way interactions. When the variances for the types A and B syllables were compared with those of the type C syllables a non-significant tendency toward lower variance on the rhythmic syllables was noted.

Question 5 of Chapter I can now be answered. Tapping and click placing seem to be two very different tasks, at least from the point of view of the resulting variances. In the tapping task, variances differ in absolute magnitude from subject to subject, probably reflecting differences in native motor-rhythmic abilities and styles of tapping. The relative magnitude of a subject's tapping variability is further a function of the type of syllable to which he is tapping. Agreement between the subjects in the ordering of syllable variances is seen from Table 13, which gives the correlations of variance magnitudes of syllables for the three subjects on the nine utterances and for all eighty-one syllables together. The correlation is great enough to indicate significant agreement between the subjects, but small enough to indicate considerable differences in the subjects' reactions. It would be too great a step to take here, however, to conclude from these differences in reactions that the subjects perceive rhythms differently. The dependence of variance size on the interaction of the subject and the syllable type may be a function of tapping styles. The subject who had the smallest overall range of variance magnitude tapped in the same manner for all syllables and showed little animation in the task. The other subjects, with comparable ranges, used other external muscle groups (head, hands, feet) to keep time with the utterance rhythm, and their tap was a part of this more complex motor behavior.

Table 13

Correlations Between Subjects on s^2, d^2, log s^2, and log d^2

	Subj. 1 vs. Subj 2		Subj. 1 vs. Subj. 3		Subj. 2 vs. Subj. 3	
	$r_{s_1^2, s_2^2}$	$r_{d_1^2, d_2^2}$	$r_{s_1^2, s_3^2}$	$r_{d_1^2, d_3^2}$	$r_{s_2^2, s_3^2}$	$r_{d_1^2, d_3^2}$
All Syllables	.40	.42	.29	.23	.36	.42
Utterance 1	.16	.20	.56	.15	.33	.38
Utterance 2	.44	.70	.17	.14	.56	.45
Utterance 3	.53	.30	.18	.22	.12	.33
Utterance 4	-.50	-.45	.71	.51	.25	.50
Utterance 5	.71	.65	.24	.35	.04	.10
Utterance 6	.22	-.07	.40	.13	.77	.87
Utterance 7	.06	.18	-.04	.16	.35	.50
Utterance 8	.45	.36	.07	-.01	.43	.51
Utterance 9	.75	.62	.62	.58	.90	.92

Logarithmic

	$r_{\log s_1^2, \log s_2^2}$	$r_{\log d_1^2, \log d_2^2}$	$r_{\log s_1^2, \log s_3^2}$	$r_{\log d_1^2, \log d_3^2}$	$r_{\log s_2^2, \log s_3^2}$	$r_{\log d_2^2, \log d_3^2}$
All Syllables	.51	.45	.39	.33	.41	.51
Utterance 1	.41	.39	.67	.39	.41	.46
Utterance 2	.65	.76	.37	.34	.49	.59
Utterance 3	.78	.53	.35	.45	.34	.55
Utterance 4	-.64	-.52	.53	.32	.29	.61
Utterance 5	.49	.57	.37	.47	.01	.14
Utterance 6	.36	.06	.50	.30	.86	.87
Utterance 7	.28	.31	-.09	.21	.27	.45
Utterance 8	.45	.41	.06	.07	.42	.60
Utterance 9	.77	.64	.54	.58	.81	.85

The absence of differences between the subjects and syllables in the click placing task indicates that it is unlike the tapping task in this respect, and is therefore less appropriate for measuring "rhythmicalness". The absolute magnitude of variance in the click task is approximately equal to that of the tapping task with an average over all syllables of about 1000 $msec^2$, making it about as reliable. It certainly is, however, a less valid indicator of the rhythm of an utterance.

3.4 Questions 1 through 3: Differences Between Means

Since Questions 1 through 3 of Chapter I ask about similarities in beat location, and since distribution means are the measures of this beat location, the three questions can be interpreted in the statistical domain as hypotheses about equality of distribution means. Thus, to say that two subjects tap at the "same place" on a syllable is to say that the means of the distributions of taps given by the two subjects on that syllable are equal. There are various statistical tools for testing whether or not two means are equal.

3.4.1 Question 2: Click Data

The results of an analysis of variance of the click distribution means are given in Table 14. The effect of syllables is inextricably bound up with the location of the timing pulse on the tape loop and so does not concern us here. The main effect of the subjects is non-significant. All other effects except the syllables-by-conditions interaction are significant.

Table 14

Click Placement Analysis of Variance Table

Source	Degr. Freedom	Sum of Squares	Mean Squares	F-Ratio	Probability Level
Conditions (C)	2	16600	8301	.41	> .5
Syllables (S)	29	80.05	2.76	0	> .5
Subjects (L)	1	3337	3337	2.48	∿ .1
C x S	58	79190	1365	1.01	∿ .5
C x L	2	39650	19820	14.7	< .0005
S x L	29	140300	4839	3.60	< .0005
C x S x L	58	85880	1481	1.10	> .25
Residual	720	968000	1345		

Marginals for Conditions (msec.)

	Subject 1	Subject 2	Total
Condition 1 -	588	590	589
Condition 2 -	603	581	592
Condition 3 -	578	586	582

The relative contributions to the variance of the different factors deserve attention. The largest contributor was the subject-by-conditions interaction. The relative order of magnitude of the means of the two subjects under the various conditions shows that subject 1 demonstrated a definite bias under the three conditions while subject 2 did not. Subject 1's click placements were 25 msec earlier in the syllable when the speech was heard in the left ear and the click in the right (condition 3) than when the speech was in the right ear and the click in the left (condition 2). With both speech and click in both ears (condition 1), subject 1's click placements were intermediate between the placements for conditions 2 and 3. Subject 2's click placements showed a different and weaker bias under the three conditions.

Even though the mean square error for conditions ranks next to that for L x C in size, the main effect of conditions is rendered nonsignificant because of the statistical model being used.

There was also a significant interaction between subjects and syllables. The distribution of differences between the means of the click distributions of the two subjects on the various syllables is decidedly bi-modal, but no ready interpretation of this interaction is available.

It is thus concluded from this analysis of variance of the click data that the kind of syllable a subject is listening to and the conditions under which he hears it have an effect on his location of a click to match the beat of that syllable. However, from the data at hand, the subjects do not appear to differ in their overall placement of the click. Thus, Question 2 of Chapter I is answered in the affirmative.

3.4.2 Question 1: Tapping Data

Because the distribution variances are non-homogeneous, the variation of the distribution means, related linearly to the distribution variances, are also non-homogeneous. Analysis of variance therefore cannot be used on the tapping data.

A natural choice for testing differences between means would be the t-test. However, because of the unreliability of individual subjects, both within and between experimental sessions, calculated t-scores would be too large and would lead to falsely significant findings. t-scores are therefore inappropriate for testing differerences between these means.

The hypothesis of equality can be tested from the ordering of the relative magnitudes of the distribution means. The size order of the means for the three subjects was compiled for all syllables, and the positional relationships that hold for the three subjects are given in Table 15. The chi-square test of effects in this contingency table is significant far beyond the .001 level. The interpretation of this contingency table is that subject 1 has a generally greater mean (taps later) than either of the other two subjects. The other two subjects are not significantly different. The conclusion can be drawn, then, that not all subjects tap in the same place, so Question 1 of Chapter I must be answered in the negative.

3.4.3. Question 3: Tapping versus Click Placement

Comparison of tap and click data requires separate consideration, since the data were gathered using different tape loops. In order to

Table 15

Contingency Table Showing Number of Times each Listener's Tapping Mean was Greatest, Middle, or Least in the Order of the Three Listeners' Means.

	Greatest Mean	Middle Mean	Least Mean
Listener 1	50	20	11
Listener 2	14	29	38
Listener 3	17	32	32

$$\chi^2_{4\ d.f.} = 47.3\ (p << .001)$$

Subtable for Listeners 2 and 3

	Greatest	Middle	Least
Listener 2	14	29	38
Listener 3	17	32	32

$$\chi^2_{2}\ d.f. < 2\ (p > .5)$$

compare means between the two experiments, the locations of the timing pulses on the two loops were established with respect to some well-defined event in the mingogram trace of the speech wave. A vertical line at this event, labeled "COMP LINE", is shown on the mingograms in Appendix A. The difference in time between the two timing pulse locations, relative to this well-defined event, was subtracted from one mean for comparison with the other. A 100-cycles-per-second triangular wave is displayed on the mingogram along with the speech wave and the timing pulse. Time measurements were thus accurate to about $\pm$ 1 msec; no calibration of reliability was carried out for this measurement process. The differences between tap and click distribution means are given in Table 16. A positive value in this table indicates that on that syllable the subject's tapping mean was later in time than his click placement mean. There is a difference between the two subjects in their tapping vs. click placing behavior. Subject 1 shows but two negative values; therefore, he generally tapped later than he placed the click. The average tap delay was approximately 3 csec. Subject 2 tapped generally earlier to two utterances (Numbers 2 and 7) and later on the third (Number 5), the average tap delay on utterances 2 and 7 being about 2 csec. A possible explanation of the difference in behavior of subject 2 on utterance 5 is that the syllable rate is greatest for that utterance and his particular tapping style did not adapt to this greater speed, giving a slight delay of his tap with respect to where he "really" felt the syllable beat to be.

Because of the size and consistent bias of the displacement of the subjects' click means from their tap means, it can therefore be concluded that the answer to Question 3 of Chapter I is: No, the subjects do not place the click where they tap.

Table 16

Differences Between Tap and Click Means in Milliseconds

+ ≡ tap precedes click

— ≡ tap follows click

	$m_{tap} - m_{click}$	
Utterance 2	Subject 1	Subject 2
ARE	0	33
PAR	-9	-15
TIC	-40	6
U	—	—
LY	-73	6
GOOD	-5	25
IN	33	-2
CUR	-34	32
RENTS	-48	6
Utterance 5		
WANTS	-99	-156
TO	-60	-66
BE	-9	-27
A	5	-42
PER	-52	-10
FOR	-11	-26
MER	-27	0
NOW	-1	1
Utterance 7		
LIKE	-40	-5
TO	-95	36
KNOW	-35	39
HOW	-13	-16
THEY	-74	15
CON	-26	21
STRUCT	-26	25
THESE	-42	66
THINGS	-32	35

3.5 Summary

Tapping the finger in time to a syllable beat and placing an audible click on that syllable beat are different behavioral tasks. The distributions of responses given by the subjects in the two tasks have different characteristics, the more important of which are form, location and dispersion. The distributions of taps by various subjects on various syllables appear to be roughly normal in form, whereas the click placing distributions are much flatter than normal, even U-shaped in form. Different subjects tap in different places, as evidenced by the great variability among the means of the various distributions, there was no such variability in the click placing task. Furthermore, although one subject showed a large bias in the placement of the click relative to where he tapped, somewhat constant over all syllables, the second subject's bias appeared to be a function of the utterance to which he was responding. Finally, although the tapping and click placing distributions have approximately the same average variance, the variances of the tapping distributions vary more widely than do those of the click distributions. Tapping variance was found to reflect in part the rhythmic character of the syllable.

CHAPTER IV - Applications of the Data

4.0 Introduction

4.0.1 Bias and Error

The main objective of the experimental work in this thesis is a clearer definition of the term "syllable beat". There was much variability in the definition of syllable beat between experiments, between subjects in the same experiment, and even within a given subject in a given experiment over an extended period of time. The displacement in time of the mean of a subject's responses from a hypothetical "point" to which he may be responding remains a problem. The variance of his responses also varies widely. Separating these two "error components" from each other, a subject's response can be considered to be the sum of his bias to one side or the other of the supposed target plus a random error for that response. Presumably the bias would remain fairly fixed from one response to another, over the short run. Probably it would be different for different subjects and for different syllables and might undergo slow fluctuations in size for the same subject or syllable over extended periods of time. The random error component would be peculiar to that response and the algebraic sum of the errors over all responses would be zero.

4.0.2 Tapping Data

Applying these notions to the data from the tapping experiment, let us consider the distribution of taps given by a listener on a syllable. We find a positive lag-one autocorrelation in the sequence of taps. The mean square successive difference measure of variation, d^2, is about eight-tenths as great as the sample variance, s^2. The difference s^2-d^2 is attributable to slow fluctuations in the bias of tapping. The variability with which a subject taps has d^2 as a minimum, and is a function of at least the listener and the syllable, as was seen from the analysis of variance of the tapping distribution variances (Section 3.3, above). The fluctuations in bias, not analyzed in detail, presumably result from changes in perceptual and motor activity. This question will not be treated any further in this work.

4.0.3 Click Data

The click data do not allow us to separate bias from error. In contrast with the situation for tapping where bias was treated as fixed in the short run, the successive click placements to a given syllable are far apart in time for the listener and so the bias must be considered as the sum of a fixed bias plus a random component. The algebraic sum of the random components is again zero. Although this random bias component is formally no different from the error, as discussed before, and cannot be distinguished from it in the data, bias remains conceptually different from error, since it results from asymmetries in the physio-perceptual system, while error defines the lower limit of accuracy of the system.

4.1 Bias Components

4.1.1 Syllable Tapping and Click Tapping Experiments

The experiment previously described (Section 3.4.2) in which subjects tapped to a sequence of four equally spaced clicks was intended partly as a calibration of the bias in tapping. The results indicate that bias is a function of the subject, the spacing of the clicks, and the position of the click in the short sequence. Table 17 gives the differences in msec between the actual location of the clicks and the means of 50 taps to each click. A positive value indicates that the mean preceded the click. The two subtables of Table 17 show that the three subjects are ordered according to the degree to which their taps preceded the click and that the three clicks are ordered according to the degree by which the tapping average for all subjects precedes them. Table 14 (p. 67) shows that subject 1 tended to tap earlier to syllables than subjects 2 and 3 but the latter do not differ in this respect. Table 18 shows the order of tapping to just the types A and B syllables. The means are now clearly ordered with subject 1 earlier than subject 2, and subject 2 earlier than subject 3; the subtable for subjects 2 and 3 shows variations significant at the .05 level. The ordering of the subjects' means in the tapping-to-clicks experiment agrees with the ordering obtained in tapping to rhythmic syllables.

The time between successive clicks on Loop 2 was half that of Loop 1. The subtables of Table 17 show a corresponding decrease in the amount by which subjects' taps precede the clicks.

The question naturally arises as to whether the bias in tapping to clicks is the same as the bias in tapping to syllables. The speech signal is a different stimulus from a sequence of evenly spaced clicks, and tapping behavior may also change between the two situations. The comparability of the tasks is shown by the similarity of form of the distributions of responses, i.e., the distributions have the same general shape and variances of comparable size, and also their means retain the same size ordering relation between subjects.

TABLE 17

Differences Between Click Location and Mean of

Fifty Taps to the Click (Milliseconds $\pm$ 1 msec).

		Loop 1			Loop 2		
Subject	Session	Click 2	Click 3	Click 4	Click 2	Click 3	Click 4
1	I	30	9	0	-2	-6	-19
	II	16	24	9	0	19	6
2	I	36	9	35	20	-7	15
	II	16	12	7	2	-7	11
3	I	60	76	61	21	21	12
	II	90	50	53	12	19	12

Subtables:

17A Average over Clicks and Sessions

Subject	Loop 1	Loop 2
1	15	0
2	19	6
3	65	16

17B Average over Subjects and Sessions

Click	Loop 1	Loop 2
2	41	9
3	30	6
4	28	6

TABLE 18

Contingency Table Showing Number of Times each Listener's Tapping Mean was Greatest, Middle or Least in the Order of the Three Listeners' Means for Types A and B Syllables Only

(Compare with Table 14)

Listener	Greatest Mean	Middle Mean	Least Mean
1	27	6	4
2	4	21	12
3	6	10	21

$$\chi^2_{d.f. = 4} = 47.8, \ p << .001$$

The biases in the tapping-to-clicks experiment are ordered according to the particular click being tapped to, but the variation of these biases is not so great as that for the two loops compared. This result suggests that bias is fairly constant for a given rate of presentation of the rhythmic stimulus. If the bias is assumed to be constant and variances within and between utterances are computed for the differences between tap and click placements in the two experiments, a significant change in bias is discovered for the two subjects over the three utterances. However, subject 1's click placements precede his taps by the greatest amount on utterance #7, with #5 having the least difference, while subject 2 has the opposite ordering, with his click placements on utterances #2 and #7 succeeding his taps. This interaction of biases presents a very confusing picture and obscures the location of the syllable beat.

It was mentioned above that $s^2 - d^2$ represents the amount of variance attributable to short-range changes in bias in the tapping experiment. Another measure of the same short range bias variability is the variance of the means of the first and second halves of the distributions. Because the last tap was given only four minutes after the first one, bias did not have a chance to change very much, and the changes must have been a function of short term effects in the tapping task. Such short term effects might have been due to changes in posture or some kind of satiation resulting from so much repetition of the utterance. Long range changes in bias are measured by the click tapping experiment, in which the subjects tapped to the same stimulus in a second session held several days after the first. Much greater changes in activity and perception result in greater differences in bias. These two measures of bias variability are compared with pooled distribution variance in Table 19.

TABLE 19

Within and Between Variance Estimates in the

Two Tapping Experiments (Milliseconds^2)

Within Variances

Subject	Tapping to Speech A and B Syllables	Tapping to Clicks Loop 1	 Loop 2
1	1227	865	370
2	741	784	341
3	2243	1561	806

Between Variances

	Tapping to Speech	Tapping to Clicks			
	1st Half vs. 2nd half of Distributions	1st Half vs. 2nd Half of Distributions		Session I vs. Session II	
Subject		Loop 1	Loop 2	Loop 1	Loop 2
1	2618	1716	1562	4182	10450
2	1835	3371	392	9942	2825
3	5383	11044	1438	13650	7082

4.1.2 Click Placing Experiment

The click data are much harder to interpret in the light of the question of constant bias. There is strong evidence for the existence of bias in subject 1's click placements, for they usually preceded the tap means, which, in turn, probably preceded the rhythmic stimulus. Subject 2's placements were less clearly to one side or the other of his taps. The above comparison is valid only if the rhythmic stimulus was the same for both the tapping and click placing experiments. It is impossible to test this hypothesis in any coherent fashion with the present data. There were too few subjects and too few observations per subject.

Subject 1 showed a consistent bias according to the condition of presentation of the speech stimulus. If he heard the speech in the right ear and the click in the left ear (condition 2), his placements were on the average later than if he heard the speech in the left ear and the click in the right (condition 3). His placements for condition 1, where he heard both speech and click in both ears, were on the average between those for conditions 2 and 3. Subject 2 showed a different and weaker bias. The number of times a subjects mean placement for a given condition was earlier or later than the placements for the other conditions is summarized for all syllables in Table 20.

4.2 Error

4.2.1 Error and Syllable Type; Tapping Data

Because bias is so difficult to calibrate, the precise location of the syllable beat is impossible to determine with the present data.

TABLE 20

Contingency Tables Showing Order of Magnitude of Each Subject's Mean Click Placement for the Three Conditions of Stimulus Presentation

Subject 1

Condition	Greatest Mean	Middle Mean	Least Mean
1	10	14	7
2	1.5	11.5	18
3	19.5	5.5	6

$\chi^2_{d.f. = 4} = 28.0,\ p < .001$

Subject 2

Condition	Greatest Mean	Middle Mean	Least Mean
1	8	14	9
2	10	13	8
3	13	4	14

$\chi^2_{d.f. = 4} < 9.10,\ p \simeq .05$

Condition 1: Speech and Click Together in Both Ears
Condition 2: Speech in Right Ear, Click in Left Ear
Condition 3: Speech in Left Ear, Click in Right Ear

What, then, can the error component of the subject's response tell us? It was pointed out in Chapter III above that the variance of a subject's taps changes depending on what syllable he is tapping to. Syllables were typed, a priori, according to the kind of rhythmic role that the experimenter thought they served when he listened to the utterances. Syllable typology will be discussed in more detail in Section 4.4, below. For this discussion of error the important aspect of syllable type is that a subject's taps had least variability on the syllables in which we are most interested, i.e., the rhythmic ones. From this fact we can draw two conclusions. First, if tapping bias can be calibrated, then the syllable beat can be more accurately pinpointed because the minimum error of tapping is less. Second, and basically more important for this whole area of research, tapping is a valid response to rhythmicalness of syllables. The actual variation of tapping variance over the several syllable types will be discussed in Section 4.4.

Sample variance, s^2, and mean square successive difference, d^2, the two measures of internal variation used in this study, are compared for all subjects and all syllables in Table 21.

4.2.2 Click Placing Data; Method of Limits Error

Variances of the click placing experiment are given in Table 22. One hypothesis to explain the U-shape of the distributions of click-placing is that the subjects perceive the syllable beat as an interval of time and move the click gradually into this interval, stopping before the click reaches the center. This hypothesis was tested by observing how often the final click placement for a trial was on the same side

TABLE 21

Syllable	s^2	d^2	d^2/s^2	Syl.	s^2	d^2	d^2/s^2	Syl.	s^2	d^2	d^2/s^2
weight	3173	2162	.68	guy	638	441	.69	how	1723	1630	.95
of	5052	2615	.52	the	701	578	.82	they	2318	2144	.92
the	2565	2012	.78	play	2532	1635	.65	con	1139	1043	.92
line	397	389	.98	was	895	756	.84	struct	1559	1392	.89
to	1059	967	.91	o	1121	1083	.97	these	572	542	.95
get	5023	3803	.76	ver	996	725	.73	things	557	490	.88
more 1	298	229	.77	wants	1448	1019	.70	like	587	517	.88
and	1586	719	.45	to	2300	1852	.81	pre 1	553	423	.76
more 2	611	459	.75	be	2278	1516	.67	dict 1	527	442	.84
out	829	505	.61	a	4858	3422	.70	ing	1024	787	.77
are	1847	1566	.85	per	1811	1027	.57	I 2	2202	931	.42
par	10815	3921	.36	for	1478	1009	.68	love	719	458	.64
tic	1254	1366	1.09	mer	1032	724	.70	to	858	482	.56
u	4826	3895	.81	now	1832	525	.29	pre 2	3162	3120	.99
ly	2568	1915	.75	terms	994	613	.62	dict 2	1117	629	.56
good	942	647	.69	of	1706	1195	.70	things	525	386	.73
in	3923	2987	.76	get	648	527	.81	hap	2199	2491	1.13
cur	1044	877	.84	ting	1853	1115	.60	pen	3275	2336	.71
rents	2176	1441	.66	a	1087	608	.56	now	637	521	.82
he	2082	1754	.84	name	736	573	.78	and	1481	822	.55
reared	419	409	.98	for	1199	1001	.84	sort	4422	2745	.62
back	493	358	.73	him	1515	1278	.84	of	3039	3056	1.01
and	2600	2110	.81	self	2261	2478	1.10	go	357	349	.98
fired	494	500	1.01	so	1610	1032	.64	out	1340	1211	.90
it	1038	772	.74	like	685	547	.80	on	2700	1105	.41
by	418	393	.94	to	959	884	.92	a	5545	3660	.66
a	1068	862	.81	know	2115	2184	1.03	limb	812	705	.87

s^2 and d^2 for All Subjects on All Syllables
(Subject 1)

TABLE 21

Syllable	s^2	d^2	d^2/s^2	Syl.	s^2	d^2	d^2/s^2	Syl.	s^2	d^2	d^2/s^2
weight	981	915	.93	guy	649	452	.70	how	978	659	.67
of	1505	1316	.87	the	1371	1151	.84	they	1161	1021	.88
the	5628	4073	.72	play	614	602	.98	con	1397	980	.70
line	665	355	.53	was	700	625	.89	struct	802	760	.95
to	1001	972	.97	o	721	536	.74	these	906	714	.79
get	1086	791	.73	ver	631	516	.82	things	517	544	1.05
more 1	1049	1139	1.09	wants	363	324	.89	like	688	824	1.20
and	3317	2191	.66	to	721	528	.73	pre 1	3710	2500	.67
more 2	998	480	.48	be	1171	977	.83	dict 1	481	475	.99
out	521	459	.88	a	4553	2728	.60	ing	566	516	.91
are	1357	866	.64	per	3541	2208	.62	I 2	1619	1261	.78
par	1985	1036	.52	for	771	506	.66	Love	639	749	1.17
tic	562	411	.73	mer	1423	1097	.77	to	904	487	.54
u	2843	1986	.70	now	337	262	.78	pre 2	2998	1677	.56
ly	2931	1730	.59	terms	612	527	.86	dict 2	491	464	.94
good	1353	326	.24	of	1791	1466	.82	things	571	505	.88
in	2285	1880	.82	get	1019	1097	1.08	hap	915	645	.70
cur	949	883	.93	ting	1041	770	.74	pen	2734	1938	.71
rents	1007	870	.86	a	1492	1113	.75	now	433	328	.76
he	1718	844	.49	name	634	502	.79	and	594	468	.79
reared	484	468	.97	for	1539	819	.53	sort	1162	813	.70
back	687	725	1.05	him	2933	1261	.43	of	632	441	.70
and	1165	786	.67	self	885	600	.68	go	414	413	1.00
fired	422	282	.67	so	1375	1135	.83	out	1360	973	.71
it	969	786	.81	like	1244	955	.77	on	1457	1078	.74
by	412	349	.85	to	2223	1455	.65	a	3175	2953	.93
a	2473	1872	.76	know	1216	1085	.89	limb	655	368	.56

s^2 and d^2 for All Subjects on All Syllables
(Subject 2)

TABLE 21

Syllable	s^2	d^2	d^2/s^2	Syl.	s^2	d^2	d^2/s^2	Syl.	s^2	d^2	d^2/s^2
weight	4243	4283	1.01	guy	1628	1555	.96	how	1679	1522	.91
of	3138	2548	.81	the	3376	3108	.92	they	1138	1021	.90
the	3882	3407	.88	play	4358	3887	.89	con	2285	1337	.59
line	2822	2028	.72	was	1976	1608	.81	struct	2413	1440	.60
to	4159	4297	1.03	o	2090	1264	.61	these	2334	1988	.85
get	3674	947	.26	ver	1748	972	.56	things	2116	1105	.52
more 1	889	918	1.03	wants	8564	6350	.74	like	3469	3570	1.03
and	2841	2655	.93	to	2907	3073	1.06	pre 1	3501	2947	.84
more 2	1551	1414	.91	be	2904	1420	.49	dict 1	1892	1900	1.00
out	1023	744	.73	a	5086	4147	.82	ing	1860	1688	.91
are	2920	2631	.90	per	2356	2383	1.01	I 2	3276	2222	.68
par	3349	2909	.87	for	1224	1170	.96	love	3385	2653	.78
tic	1674	1441	.86	mer	1928	1870	.97	to	1093	677	.62
u	3134	2634	.84	now	1942	1317	.68	pre 2	2433	2017	.83
ly	5872	6147	1.05	terms	1177	1008	.86	dict 2	1805	1246	.69
good	1187	947	.80	of	5450	4676	.86	things	1719	1336	.78
in	2125	1334	.63	get	1995	1676	.84	hap	1572	1256	.80
cur	3928	2989	.76	ting	2682	2538	.95	pen	3374	2716	.81
rents	2188	1614	.74	a	2670	2443	.91	now	1511	1152	.76
he	2619	2638	1.01	name	1213	844	.70	and	1304	853	.65
reared	2702	1656	.61	for	1885	1136	.60	sort	1316	1059	.80
back	2326	1634	.70	him	4331	3164	.73	of	2157	1764	.82
and	2250	1668	.74	self	1798	1386	.77	go	1144	923	.81
fired	2050	1590	.78	so	3837	2894	.75	out	2440	2041	.84
it	4777	4422	.93	like	1903	980	.51	on	2323	2363	1.02
by	1053	774	.73	to	2844	2248	.79	a	4439	3693	.83
a	2060	2216	1.08	know	3023	2752	.91	limb	2329	1191	.51

s^2 and d^2 for All Subjects on All Syllables
(Subject 3)

TABLE 22

Subject 1

Syllable	Condition 1	Cond 2	Cond 3	Total	Syl	Cond 1	Cond 2	Cond 3	Total
spin	1069	321	126	617	per	2474	1019	3315	2355
ers	1038	594	1114	862	for	404	510	1748	1264
are	2504	1240	939	1933	mer	2569	530	1211	1960
par	928	1011	814	902	Now	1627	1584	3411	1974
tic	1259	1003	429	822	I	195	955	390	710
ly	2394	2795	1904	2326	would	1489	2162	2679	2478
good	552	2376	329	1032	like	210	943	1217	992
in	1754	996	1278	1353	to	3849	1526	3557	3280
cur	2220	1990	4155	2413	know	324	1243	307	814
rents	202	1212	2918	1804	how	282	3219	465	1571
He	1466	421	786	964	they	943	3062	2012	2122
Wants	368	509	56	1081	con	2179	334	402	950
to	334	1212	3370	1499	struct	166	2036	1614	1174
be	2989	827	625	1747	these	743	936	109	728
a	187	4392	2069	1956	things	1604	4498	671	1228

Subject 2

Syllable	Condition 1	Cond 2	Cond 3	Total	Syl	Cond 1	Cond 2	Cond 3	Total
Spin	762	1660	1775	1203	per	306	2254	860	1102
ners	688	348	214	384	for	2439	134	1024	1103
are	756	372	2047	1158	mer	1603	1728	360	1434
par	2022	2444	933	2063	now	353	1173	1877	973
tic	1622	1351	938	1157	I	692	1330	1240	982
ly	1372	304	4399	1967	would	2549	462	3057	1878
Goodd	3906	892	770	1773	like	137	734	907	587
in	1181	3090	860	1663	to	1478	585	1312	1053
cur	1191	3752	428	1726	know	2680	1054	1627	1549
rents	1688	1552	824	1239	how	2070	501	1361	1750
he	1868	1060	178	1192	they	1040	229	1094	887
wants	498	882	432	167[illegible]	cpn	1380	1391	1732	1336
to	2345	1808	2141	1986	struct	3347	956	471	1671
be	595	2503	288	1299	these	311	1665	3385	2055
a	1812	3909	992	1948	things	288	1890	522	986

Condition 1: Speech and Click Together in Both Ears
Condition 2: Speech in Right Ear, Click in Left Ear
Condition 3: Speech in Left Ear, Click in Right Ear

s^2 For Conditions is Computed over the Five Observations for that Condition on that Syllable by that Subject

s^2 For Total is Computed over the Fifteen Observations for that Syllable by that Subject

s^2 for Three Conditions and Grouped Conditions in Click Placing Experiment of the mean placement for the subject as the direction from which the click last came. If the subject were stopping short of the center of the assumed "beat interval," then when the click moved in last from the early side, there would be more instances in which click placement preceded the mean and when it came from the late side, more instances in which the click placement was after the mean. The results of this test are given in Table 23. They show that both subjects tended to stop short of the mean under both conditions of click movement. This tendency to stop short of the center of an interval-like stimulus is known as bias in the method of limits (Smith, 1957). Because of the known effect of condition of stimulus presentation, each placement was compared with the mean placement for the condition under which that placement was made.

4.3 Phonetic Applications

4.3.1 Location of the Syllable Beat

Miyake (1902), Hollister (1937), and Classe (1939) single out the onset of the nuclear vowel as the rhythmic maximum of the syllable beat. Because this point in the physiologic-acoustic speech sequence is associated with many motor activities and perceptual cues, it is a natural choice for such a maximum. There are generally large articulator movements as the last consonant before the nuclear vowel is released. In stressed syllables a chest pulse often accompanies the onset of the vowel. If the initial consonants are voiceless, the onset of voicing requires the appropriate tensing of the vocal folds and the maintenance of sufficient air flow through the glottis. Thus, movements in the

TABLE 23

Contingency Table Showing Bias in the Method of Limits for the Click Placing Experiment

Subject 1		Side of Final Click Placement: Above	Below	Total
Direction from which click last came	Above	119	113	232
	Below	99.5	133.5	233
	Total	218.5	246.5	465

$\chi^2_3 = 3.44$ N.S.

Subject 2		Side of Final Click Placement: Above	Below	Total
Direction from which click last came	Above	129	111	240
	Below	100	125	225
	Total	229	236	465

$\chi^2_3 = 4.17$ N.S.

Subjects 1 & 2 together		Final Placement: Above	Below	Total
Direction	Above	248	224	472
	Below	199.5	258.5	458
	Total	447.5	482.5	930

$\chi^2_3 = 7.52$ $p \sim .06$

Comparison of Subjects 1 & 2 yields $\chi^2_{d.f. = 3} = .35$, $p > .5$

articulators, chest wall, and larynx are the source of many kinesthetic stimuli associated with the onset of the nuclear vowel.

The acoustic wave shows a great increase in energy and striking changes in spectral structure at the beginning of the nuclear vowel. As will be pointed out later, these changes occur to different degrees and at different rates for different types of consonantal release, but they are strong perceptual cues in all cases. The vowel onset is therefore a clearly distinguished event for the listener as well as for the speaker.

Table 24 gives the amount by which the three subjects' tapping means precede the onset of the nuclear vowel in the rhythmic (types A & B) syllables.[6] The average values are quite well matched with each subjects' precession in tapping to clicks for loop 1 with the more widely spaced clicks. This agreement suggests that subjects react in the tapping experiment as though they were tapping to a sharply defined stimulus occurring at the time of onset of the nuclear vowel of the stressed syllable. An assumption underlying this conclusion is that all syllables are groupable, even though we have seen that bias changed in the click tapping experiment depending on the click to which the subject was tapping.

Newcomb (1960, 1961) gives a more complex rule for beat location (see Section 1.2.5, above), involving the consonant sequences that precede the nuclear vowel. The click location means in his experiments were: (1) "at the onset of voicing" for voiceless obstruent consonants;

---

[6]Measurements were derived from spectrograms made on a BTL model D spectrograph and from mingograms showing simultaneous tracings of the speech wave, speech power, and a 100 cps triangular timing signal. Measurements were taken to the nearest .01 second, as finer measurements were not thought to be justified.

TABLF 24

Amounts by which Tapping Means Precede Nuclear Vowel Onset

Times in Selected Type A and B Syllables (Milliseconds ± 10 msec)

Syllable	Subject 1	Subject 2	Subject 3	Syllable	Subject 1	Subject 2	Subject 3
Weight	0	6	20	Self	60	72	112
Line	-11	20	10	So	31	67	135
More 1	-17	23	46	Like	-44	18	12
More 2	76	80	94	Know	41	108	94
Out	4	26	21	How	18	35	59
Tic	36	97	76	Struct	51	73	115
Good	6	79	58	These	-3	68	134
Cur	41	97	157	Things	12	53	96
Play	131	22	170	Dict 1	-21	-15	-12
Was	22	15	50	Love	-1	2	14
O	-50	-32	9	Dict 2	5	13	-22
Ver	-4	-14	30	Things	48	32	35
Be	28	39	33	Hap	18	-7	-27
For	53	10	29	Now	-32	9	40
Now	38	39	24	Sort	60	45	61
Terms	9	58	33	Go	24	32	19
Get	32	38	40	On	45	-4	42
Name	23	-8	-12	Limb	-24	49	64

Mean Precession Over All Syllables

Subject 1	20 msec
Subject 2	35 msec
Subject 3	52 msec

Mean Precession of Tap to Click in Loop 1

Subject 1	15 msec	(From Table 17A)
Subject 2	19 msec	
Subject 3	65 msec	

(2) "at the release of consonantal articulation" for voiced obstruents; and (3) "at the beginning of the return from the extreme point of formant deflection toward the position of the following vowel" for semivowels (Newcomb, 1961, p. 3). In the case of voiceless obstruent consonants, the onset of voicing is the same as the onset of the nuclear vowel, and so the same kinesthetic and acoustic cues are available. In the case of voiced obstruents, laryngeal tensions and air flow are already sufficient for voicing, and so the significant motor activities are probably articulator and chest wall movements. The rise in acoustic energy is great but not as great as for voiceless obstruents and the change in spectral structure is not as great nor as abrupt as in the change from no-voicing to voicing. In semivowels, fricatives, and sibilants there is no sudden articulator movement, but rather a smooth motion through a point of maximum, but incomplete, closure. This maximum of tension just before the consonantal release, along with chest wall movements, is a possible kinesthetic cue for the rhythmic beat. The point of maximum articulator tension is probably also the point of maximum deflection of the formants of the semivowel, and the turning of these formants toward the nuclear vowel position plus the associated rise in acoustic energy would be salient perceptual cues.

It is unfortunate that the present data do not permit a test of Newcomb's rule. The tapping means, which probably precede the perceptual beat, more often than not come after the point at which he would locate the beat. This precession of the beat by his location would not be critical provided that the bias could be calibrated. Since the two subjects showed opposite tendencies in their biases, this calibration is impossible. A test of Newcomb's rule would require a study with more subjects.

If we group the various syllables according to the types of consonant sequence that preceded the nuclear vowels, as in Table 25, some agreement among the subjects is found on the degree of precession of the vowel onset by the tap means. The initial consonant types in the syllables where all the subjects tapped earlier than their mean precession are semivowel (/w/ in this case), sonorant (/l/ in this case), and open (no initial consonant). The types of consonants that preceded nuclear vowels of syllables to which the subjects tapped later are fricative, sibilant, and sequences of two or more consonants (/pl/ and /str/). Stop and nasal consonants yielded mixed results.

The possibility that the duration of the sequence of the initial consonants may determine the degree of precession of the subjects' taps was tested by correlating the two sets of times. Correlations of .66, .20, and .33 were obtained in this way for the three subjects. The measured durations of the sequences of initial consonants are given in Table 26, along with the tapping precession and the derived correlations. These modest correlations indicate that the onset of the nuclear vowel may be taken roughly as the beat of the syllable, and the time by which a subject's taps precede the vowel onset will be a function of the subject's individual bias and the length or phonetic type of the initial consonant (since consonant type and consonant length are highly correlated). The measurements of a subject's bias, vowel onset, and consonant length are too crude to make any more definite statement.

4.3.2 Error in Location of the Syllable Beat

The lengths of the initial consonants, given in Table 26, range from 0 to 140 msec, with an average of 75 msec per phoneme. This value is

TABLE 25

Mean Amount of Precession of Vowel Onset by Tap Mean in Selected Type A and B Syllables Grouped According to Phonetic Type (Milliseconds ± 10 msec)

Phonetic Type	Subject 1	Subject 2	Subject 3
Open	0	-3	24
Stop	18	49	43
Semivowel (/w/)	11	10	35
Sonorant (/1/)	-20	22	25
Nasal	21	42	48
Fricative + Sibilant	30	42	75
Cluster	91	48	142
Mean Over All Types	20	35	52

TABLE 26

Initial Consonant Lengths and Correlation of These Lengths with Amount by Which Tapping Mean Precedes Nuclear Vowel Onset for the Three Subjects (See Table 25 for Tapping Precession) (in Milliseconds ± 5 msec)

Syllable	Consonant Length (msec)	Syllable	Consonant Length (msec)
Weight	90	Self	110
Line	100	So	170
More 1	105	Like	25
More 2	100	Know	80
Out	0	How	105
Tic	40	Struct	130
Good	60	These	80
Cur	140	Things	40
Play	120	Dict 1	80
Was	60	Love	100
O	0	Dict 2	75
Ver	85	Things	140
Be	30	Hap	130
For	130	Now	80
Now	70	Sort	80
Terms	40	Go	60
Get	40	On	0
Name	65	Limb	90

not out of line with the average value of 50 msec per phoneme in connected speech as suggested by Joos (Shen & Peterson, 1962, p. 12). The standard error in both tapping and click placing is approximately 30 msec, and so the distributions could be associated with one or perhaps two phonemes. Different subjects, however, have different biases, and so the correction for bias must be included before the distributions may be compared with phonetic events. Table 27 gives the average differences between and within subjects for the tapping means and between both the subjects and the conditions in the click placing means. The <u>average absolute</u> difference is the average of the unsigned amount by which two means differed and gives an indication of the expected distance between two means. The <u>average</u> difference, the difference between the biases of the two means, is an overall correction term for comparison of locations. The absolute differences add considerably to the error in location of the mean. Adding the 30 msec standard error of tapping on click placement to the, say, 10-20 msec minimum error in subject bias, the total error has approximately the length of a phoneme. Observation of many responses can reduce the standard error associated with tapping and click placing, but new experimental techniques would be required to reduce the 10-20 msec error in bias.

The phonetic events that Miyake, Hollister, Classe, and Newcomb associate with the rhythmic beat are more accurately located by standard spectrographic techniques. The error in such a location would be on the order of one or two periods of the fundamental frequency, or about a csec. Therefore, precise correlation of behavioral responses with physiologic-acoustic events must await more accurate calibration of

TABLE 27

Differences Between Mean Tapping and

Click Placing Locations, in Milliseconds

I - Differences Between Tapping Means

	Subj 1-Subj 2	Subj 1-Subj 3	Subj 2-Subj 3
Average Absolute	30	31	40
Average	15	33	17

II - Difference Between 1st and 2nd Half Means of Tapping Distributions

	Subject 1	Subject 2	Subject 3
Average Absolute	10	9	18

III - Difference Between Click Placing Means

	Subject 1-Subject 2
Average Absolute	22
Average	14

IV - Difference Between Means for Conditions of Speech-Click Presentation

	Subject 1			Subject 2		
	Cond 1-Cond 2	C1-C3	C2-C3	C1-C2	C1-C3	C2-C3
Average Absolute	19	29	20	16	19	24
Average	-1	16	17	-4	6	12

bias in response. Improved experimental techniques will probably contribute greater accuracy than increases in the amount of data collected.

4.4 Syllable Type

After the speech materials for the two experiments had been selected, but before any testing was done, the rhythms of the utterances were examined intuitively and each syllable was assigned a type according to the role the experimenter thought the syllable played in the rhythm. Assignments were made by the experimenter after listening several times to each utterance. The three general syllable types were: stressed, and therefore rhythmically accented (type A); reduced stressed or unstressed, but still rhythmically accented (type B); and unstressed, unrhythmically accented (type C). There were two sub-classes of type B syllables, corresponding to the degree of rhythmic accent. Four sub-classes of the type C class corresponded to syllables which immediately succeeded or preceded stressed syllables, or both, or neither. This syllable typology is summarized in Table 28, in which are shown the syllables of all utterances and their assigned types. Two types were assigned to some syllables in ambiguous cases. For purposes of data analysis, the first of the two ambiguous types was chosen.

Syllables were typed in this way because such an *a priori* specification of utterance rhythm permits comparison of the experimental results with native speakers' intuitions about rhythm. After the three subjects had tapped to all syllables, one extra session was devoted to the investigation of their intuitions. Each subject listened to each utterance and described it rhythmically in whatever rhythmic notation

TABLE 28

Syllable Types for All Syllables of All Utterances
(Marked Over Orthographic Vowel of the Syllable)

	Syllable types	Utterance
Utterance 1	A C2 A C1 C2 A B2 B2 A C1 B2 A	Use the weight of the line to get $more_1$ and $more_2$ out
Utterance 2	A C1 C3 C2 A C1 (C1-C3) B1 C3 A C1	Spinners are particu(lar)ly good in currents
Utterance 3	B2 A B2 A A (B2-C1) A C1 A C12 A	See when he reared back and fired it by a guy
Utterance 4	A C1 C2 A B2 A B1	After the play was over
Utterance 5	C2 A (C2-C12) B1 C1 C2 A C1 B1	He wants to be a performer now
Utterance 6	A C12 A C12 A C1 C2 A C3 C2 B1 B1	Talks in terms of getting a name for himself so
Utterance 7	A C12 B1 C2 B1 A (B2-C1)(B2-C2)A B1 B1	I would like to know how they construct these things
Utterance 8	C2 A C2 A C1 C2 A (C1-B2)(C2-B2)A B1	I_1 like pre_1dict_1ing I_2 love to pre_2dict_2 things
Utterance 9	A C1 B1 C2 A C3 C3 C2 A C1 B1 C2 A	What will happen now and sort of go out on a limb

Type A - Primary Rhythmic Accent

B1 - Reduced Stress but Major Rhythmic Beat

B2 - Reduced or Unstressed, Counterpoint in Rhythm

C1 - Unstressed, not in Rhythm, Follows Type A Syllable

C2 - Unstressed, not in Rhythm, Precedes Type A Syllable

C12- Unstressed, not in Rhythm, Between Two Type A Syllables

C3 - Unstressed, not in Rhythm, Other Than C1, C2 or C12.

he wished to use. Subjects were asked to comment on the "strictness" of whatever rhythm they perceived, this strictness presumably meaning adherence to equality of time intervals between beats. There was by no means complete agreement among the subjects as to rhythmic characterization. The subjects agreed fairly well on the rhythmic accent of any given syllable. Eight syllables marked by the experimenter as accented were not thus marked by any subject. Seven of these were marked type "B2"; the other was the "B1" syllable "like" in utterance #7. Three syllables marked as accented by at least one subject were considered by the experimenter to be unaccented. These were "are" of utterance #2 (type C2), and "I_1" and "to" of utterance #8 (types C2 and C1-B2, resp.). This finding would indicate that the B2 type is not valid, since seven of the nine B2 syllables were considered to be unaccented by all three subjects.

A more careful experiment was carried out using linguistically trained subjects. Five subjects with varying degrees of linguistic training were asked to transcribe the nine stimulus utterances in two separate sessions a week apart. (Instructions to the subjects are given in Appendix D.) Primary attention was given to the prosodies in these transcriptions. Table 29 gives the nine utterances with the stress markings by the five subjects in each session. This table is condensed into Table 30 by giving one point to a syllable for each primary or secondary stress marked on that syllable by a subject. Thus a syllable would be given a score of ten if all five subjects marked that syllable as stressed in both sessions. A zero score would result from no subject marking the syllable as stressed in either session. One subject used a four-level

TABLE 29

Subject	SYLLABLE	Use	the	weight	of	the	line	to	get	more₁	and	more₂	out
	Syllable type	A	C2	A	C1	C2	A	B2	B2	A	C1	B2	A
	Session 1	|		|			O|			O|			|
1	Session 2	|		|			O|			|			O
	" 1	/					\			\			/
2	" 2	\		\			/			\			\
	" 1	/					/						/
3	" 2	\					/						\
	" 1	/		\			/						/
4	" 2	/					/						/
	" 1	∧	∪	\	∪	∪	/	∪	\	∧	∪	\	/
5	" 2	∧	∪	∧	∪	∪	/	∪	\	∧	∪	\	/

Subject		Spin	ners	are	par	tic	u	ly	good	in	cur	rents
	Syllable type	A	C1	C3	C2	A	C1	C1-C3	B1	C3	A	C1
	Session 1	|				|					O|	
1	" 2	|				||					O||	
	" 1	\				\			\		/	
2	" 2	\				\					/	
	" 1	/									/	
3	" 2	/				\					/	
	" 1	/				/					/	
4	" 2	/				/					/	
	" 1	∧	\	∪	∪	∧	∪	∪	\	∪	/	∪
5	" 2	∧		∪	∪	//∧	∪	∪	\	∪	/	∪

Subject	SYLLABLE	See	when	he	reared	back	and	fired	it	by	a	guy
	Syllable type	B2	A	B2	A	A	B2-C1	A	C1	A	C12	A
	Session 1				|O			O		O		O
1	" 2				||			||		||		|
	" 1	\	\		\	\		/		\		\
2	" 2	\	\		\	\		/		\		\
	" 1				/			/				\
3	" 2				\			/				\
	" 1							/				
4	" 2	/	/		/	/		/		/		/
	" 1	\	∧	∪	∧	∧	∪	/∧	∪	/	∪	\
5	" 2	\	∧	\	∧	\∧	∪	/		/	∪	\

Stress Markings by All Subjects on All
Syllables in Both Experimental Sessions

(Utterances 1-3)

TABLE 29

Subject	SYLLABLE	Af	ter	the	play	was	o	ver
	Syllable type	A	C1	C2	A	B2	A	B1
	Session 1	|			|		OI	
1	" 2	|			|		|	
	" 1	\			\		/	
2	" 2	\			\		/	
	" 1	/					/	
3	" 2	/					\	
	" 1	/					/	
4	" 2	/	/				/	
	" 1	∧	∨	∨	∧	∨	/	\
5	" 2	∧	∨	∨	∧	∨	/	∧

Subject	SYLLABLE	He	wants	to	be	a	per	for	mer	now
	Syllable type	C2	A	C2-12	B1	C1	C2	A	C1	B1
	Session 1		|					OI		
1	" 2		|					OI		
	" 1		\					/		\
2	" 2		\					/		\
	" 1		/		\			/		
3	" 2		/		\			\		
	" 1		/					/		
4	" 2		/					/		
	" 1		∧	∨	\	∨	∨	/	\	∧
5	" 2	\	∧	∨	\	∨	∨	/	∨	\

Subject	SYLLABLE	Talks	in	terms	of	get	ting	a	name	for	him	self	so
	Syllable type	A	C12	A	C12	A	C1	C2	A	C3	C2	B1	B1
	Session 1	|		|		|			OI				
1	" 2	|		|		|			OI				
	" 1	\		\		\			/			\	\
2	" 2	\		\					/			\	\
	" 1	/							/				
3	" 2	/							\				
	" 1	/							/				
4	" 2	/		/					/				/
	" 1	∧	∨	∧	∨	∧	∨	∨	/	∨	\∨	∧	∧
5	" 2	∧	∨	∧	∨	∧	∨	∨	/	∨	∨	∧	\

Subject	SYLLABLE	I	would	like	to	know	how	they	con	struct	these	things	
	Syllable type	A	C12	B1	C2	B1	A	B2-C1	B2-C2	A	B1	B1	
	Session 1	OI				|				OI			
1	" 2	||								OI			
	" 1	/		\		\	\			/			
2	" 2	/		\		\				/	\	\	
	" 1	/								/			
3	" 2	/								\			
	" 1	/								/			
4	" 2	/	/							/			
	" 1	/	∨	\	∨	\	∧		∨	/	\	\	
5	" 2	/	∨	∧	∨	∧\	∧	\	∨	/	\	\∧	

Stress Markings by All Subjects on All Syllables in Both Experimental Sessions

(Utterances 4-7)

TABLE 29

Subject	SYLLABLE	I_1	like	pre_1	$dict_1$	ing	I_2	love	to	pre_2	$dict_2$	things
	Syllable type	C2	A	C2	A	C1	C2	A	C1-B2	C2-B2	A	B1
	Session 1		I		OI			OI			I	
1	" 2		OI		I			II			OI	
	" 1		/		\			/			\	\
2	" 2		/		\			/			?	
	" 1		\		/			/				\
3	" 2		/		\			/			\	
	" 1				/			/			/	
4	" 2	/			/			/			/	
	" 1	\	∧	∨	/	\	\	∧ /	∨	∨	/	\
5	" 2	\	∧	∨	/	∨	\	∧	∨	∨	/	\

Subject	SYLLABLE	What	'll	hap	pen	now	and	sort	of	go	out	on	a	limb
	Syllable type	A	C1	B1	C2	A	C3	C3	C2	A	C1	B1	C2	A
	Session 1	OI				O				OI				OI
1	2	I				OII				II				:
	1	/		\		/				/	\			/
2	2	\		\		\				/	\			/
	1			/						/				\
3	2	\		\						/				
	1	/								/				/
4	2	/				/				/	/			/
	1	∧	∨	\	∨	/	∨	∨	∨	/∧	\	\∨	∨	/
5	2	∧	∨	\	∨	/	∨	∨	∨	∧	∧\	∨	∨	/

MEANINGS OF MARKINGS

Subject 1- Three Stress System: I = Stress, II = Heavy Stress
[Sometimes Includes O= Head of Rhytm Unit (Pike)]
Subject 2- Three Stress System: / = Primary Stress
\ = Reduced Stress
Subject 3- Same as Subject 2
Subject 4- Same as Subject 2
Subject 5- Four Stress System (Trager-Smith): / = Primary,
∧ = Secondary, \ = Tertiary, ∨ = Weak
[Two Markings Under Same Syllable in Same Session Indicate Unresolved Ambiguity. Leftmost Mark Choosen for Data Analysis.]

Stress Markings by All Subjects on All Syllables in Both Experimental Sessions

(Utterances 8-9)

TABLE 30

Condensation of Table 29 by Summing Stress Marks

Syllable	Use	the	weight	of	the	line	to	get	more$_1$	and	more$_2$	out
Type	A	C2	A	C1	C2	A	B2	B2	A	C1	B2	A
Score	10	0	6	0	0	10	0	1	6	0	1	10

Syllable	Spin	ners	are	par	tic	u	ly	good	in	cur	rents
Type	A	C1	C3	C2	A	C1	C1-C3	B1	C3	A	C1
Score	10	1/2	0	0	9	0	0	3	0	10	0

Syllable	See	when	He	reared	back	and	fired	it	by	a	guy
Type	B2	A	B2	A	A	B2-C1	A	C1	A	C12	A
Score	5	5	1/2	9	5	0	10	0	7	0	9

Syllable	Af	ter	the	play	was	o	ver
Type	A	C1	C2	A	B2	A	B1
Score	10	1	0	6	0	10	2

Syllable	He	wants	to	be	a	per	for	mer	now
Type	C2	A	C2-C12	B1	C1	C2	A	C1	B1
Score	1/2	10	0	4	0	0	10	1/2	4

Syllable	Talks	in	terms	of	get	ting	a	name	for	him	self	so
Type	A	C12	A	C12	A	C1	C2	A	C3	C2	B1	B1
Score	10	0	7	0	5	0	0	10	0	1/2	4	5

Syllable	I	would	like	to	know	how	they	con	struct	these	things
Type	A	C12	B1	C2	B1	A	B2-C1	B2-C2	A	B1	B1
Score	10	1	4	0	4	3	1	0	10	2	2

Syllable	I$_1$	like	pre$_1$	dict$_1$	ing	I$_2$	love	to	pre$_2$	dict$_2$	ing
Type	C2	A	C2	A	C1	C2	A	C1-B2	C2-B2	A	B1
Score	2	10	0	10	1/2	1	10	0	0	9	4

Syllable	What	'll	hap	pen	now	and	sort	of	go	out	on	a	limb
Type	A	C1	B1	C2	A	C3	C3	C2	A	C1	B1	C2	A
Score	9	0	6	0	7	0	0	0	10	4	1/2	0	9

stress system (Trager-Smith), and half a point was given to syllables with a tertiary or weak stress marking (\`).

Table 30 is further reduced into Table 31 by tabulating the scores for the different types of syllables (experimenter's typology). The scores for type A syllables are with one exception five or greater. The scores for type B1 syllables are between two and six, again with one exception. Type B2 syllables have scores between zero and one, with one exception. Type C syllables score mostly zero, with one scoring four and one scoring two.

The experimenter's syllable typology relates strongly, then, to the degree of stress marked by trained listeners to the set of syllables.

The overall syllable typology was shown to be behaviorally valid by the analysis of variance of the tapping variances (see Section 3.3, p. 60). Types A and B syllables have generally lower variances than the type C syllables. If "average" variances are computed by averaging logarithms of the distribution variances, the type A syllables are lowest at 1048 $msec^2$, then the type B syllable, at 1485 $msec^2$, and finally the type C syllables, at 2161 $msec^2$. The type B1 syllables averaged 1328 msec, and the type B2 syllables, considered by the subjects to be unaccented and by the linguists to be less stressed, averaged 1863 $msec^2$. These average tapping variances are given in Table 32, along with the average "stressedness" score by the linguists. The variability in tapping decreased (or reliability in tapping increased) as stress increased.

The validity of tapping variance as a measure of syllable type may be due to the different kinds and amounts of rhythmic information carried by different syllables. A stressed syllable can carry the information

TABLE 31

All Scores Over Syllable Types, with Means

Syllable Type	Scores	
A	10, 6, 10, 6, 10, 10, 9, 10, 5, 10, 5, 10, 7, 9, 10, 6, 10, 10, 10, 10, 7, 5, 10, 10, 3, 10, 10, 10, 10, 9, 9, 7, 10, 9	Mean = 8.6
B1	3, 2, 4, 4, 4, 5, 4, 4, 2, 2, 4, 6, 1/2	Mean = 3.4
B2	0, 1, 1, 5, 1/2, 0, 0, 1, 0	Mean = .94
C1	0, 0, 1/2, 0, 0, 0, 0, 1, 0, 1/2, 0, 1/2, 0, 0, 4	
C2	0, 0, 0, 0, 1/2, 0, 0, 0, 1/2, 0, 2, 0, 1, 0, 0, 0, 0	
C12	0, 0, 0, 1	
C3	0, 0, 0, 0, 0	Mean = .028

TABLE 32

Average Variance _vs_ Average Stress Score

for Types A, B1, B2 and C Syllables

Syllable Type	Average Variance	Average Stress Score
A	1048 $msec^2$	8.6
B1	1328	3.4
B2	1683	.94
C	2161	.028

of a rhythmic down beat, and since people are used to tapping their fingers on down beats the task is natural and the variance small.

Syllable typology is closely related to the grammatical notion of open and closed class words in English. The types A and B syllables are likely to be the stressed syllables of open class words such as nouns, verbs, and adjectives, while the type C syllables are more likely to be closed class monosyllabic function words such as prepositions, pronouns, and conjunctions,along with the unstressed syllables of open class words. The syllables of all utterances are ranked in Table 33 according to the variability with which the subjects tapped to them. It can be seen that the function words and unstressed syllables of open class words have generally higher rankings than the stressed syllables of open class words. This rule, however, has many exceptions.

4.5 Comparison of the Two Experimental Tasks

The two experiments were designed to give two kinds of information: the location of the syllable beat, and the role of the syllable in the overall rhythm of the utterance. The important measure for locating the syllable beat is the bias of the mean of the distribution of responses; the variance of the response distribution gives information about the rhythmic role of the syllable. The tapping task and the click placing task differed in generating these two kinds of information.

Because many responses were gathered in a short period of time in the tapping experiment, tapping bias was measurable. In tapping to a sequence of clicks, the subjects showed similar tapping behavior and comparable bias. The results of these two experiments could be combined

TABLE 33

Rank of Sum of Ranks of Syllables of Stimulus Utterances, with Syllable Types

Rank of	Utterance 1		Utterance 2		Utterance 3		Utterance 4		Utterance 5		Utterance 6		Utterance 7		Utterance 8		Utterance 9	
Σ Rank	Syllable	Type	Syl	Type	Syl	Type	Syl	Type	Syl	Type	Syl	Type	Syl	Type	Syl	Type	Syl	Type
1	$more_1$	a	good	b_1	by	a	ver	b_1	for	a	terms	a	things	b_1	things	b_1	go	a
2	line	a	tic	a	fired	a	was	b_2	mer } tie	c_1	name	a	these	b_1	$dict_1$	a	now	a
3	$more_2$	b_2	cur	a	guy	a	the	c_2	now }	b_1	get	a	like	b_1	$dict_2$	a	and	c_3
4	out	a	rents	c_1	reared	a	play } tie	a	wants	a	self	b_1	how	a	to	(b_2c_1)	limb	a
5	to	b_2	are	c_3	back	a	o }	a	per	c_2	a } tie	c_2	they	(c_1b_2)	ing	c_1	hap	b_1
6	weight } tie	a	par	c_2	a	c_{12}			to } tie	(c_2c_{12})	for }	c_3	struct	a	love	a	of	c_2
7	and }	c_1	in	c_3	and	(b_2c_1)			be }	b_1	ting } tie	c_1	com	(b_2c_2)	like	a	sort	c_3
8	get	b_2	u	c_1	it	c_1			a	c_1	so }	b_1	to	c_2	pre_1	c_2	out	c_1
9	of	c_1	ly	(c_1c_3)	he	b_2					him	c_2	know	b_1	I_2	c_2	on	b_1
10	the	c_2									of	c_{12}			pre_2	(b_2c_2)	pen	c_2
11																	a	c_2

to show a moderate correlation of the mean time of a subject's tap with the time of onset of the nuclear vowel of the accented syllables. Displacement (forward) in time of the subject's taps was shown to depend on the particular subject and the length of the consonant or consonants preceding the vowel.

Bias was also in evidence in the click placing experiment, but the size of the bias did not change coherently between the two subjects. Comparison of the click placements with the speech was therefore impossible. Subjects were not significantly different in their mean click placement, although a tendency toward difference was apparent ($p \sim .1$). If, in a similar experiment with more subjects, no inter-subject differences resulted, then bias would no longer be a problem and the click placements could be compared directly with the speech.

The variability of a subjects' tapping was shown to be strongly related to the stress and rhythmic accent of the syllable, with lower variability associated with the more accented syllables. This relative size of variability correlated highly with both subjects' and trained linguists' intuitions about the accents.

The range of variability shown by subjects in the click placing experiment was much less than that in the tapping experiment. Correlations of variabilities between themselves and with other measures were therefore markedly reduced. It is doubtful that a larger experiment would increase these correlations to a size comparable with those of the tapping experiment.

A final consideration in the comparison of the two experiments is the relative ease of gathering data. In the tapping experiment, a single datum is obtained on each rotation of the loop, however long it may be. An average of six revolutions was required for each click placement, (four for one subject, eight for the other) making data six times slower to collect. A different experimental design, involving forced choice judgments of randomly placed clicks would remove this time element as well as the bias in the method of limits.

4.6 Summary and Conclusions

Two statistical measures were derived from the distributions of responses given by the subjects in the experiments described above. The means of the distributions were used to locate the syllable beat. It was found that these means are subject to biases resulting from differences between the subjects, differences between the syllables, and differences within the subjects over time. The biases were more consistent and more easily calibrated in the tapping experiment than in the click placing experiment. Simple hypotheses for the prediction of the location of the syllable beat, suggested from the literature, were tested by comparing the subjects' mean responses with these predicted locations. The displacement in time of a subject's tapping mean from the onset of the nuclear vowel of a stressed syllable was found to be moderately correlated with the length of the consonant sequence preceding the vowels.

The variances of the distributions were used as a measure of the rhythmicalness of the syllable. The tapping experiment yielded a greater range of variances than did the click placing experiment. The magnitude of the subjects' variances in tapping to a syllable was found to correlate highly with: (1) the role of the syllable in the rhythm of the utterance, according to the experimenter's and the subjects' intuitions, (2) the stress markings assigned by linguists to the syllable, and (3) the grammatical class of the word to which the syllable belongs. Specifically, syllables with lower tapping variances are more likely to be Type A or B syllables, according to the syllable typology discussed in Section 4.4, above; these low-variance syllables are more likely to be marked as stressed by trained linguists; these syllables are more likely to be the stressed syllables of open class words.

From this finding of agreement between rhythmicalness-stress and tapping behavior it can be concluded that rhythm exists in conversational English, insofar as the stimulus utterances used in this experiment are representative of conversational English. It can further be concluded that it makes sense to talk of "stress-timing" in English.

From the finding of agreement between bias in tapping to clicks and bias in tapping to the syllables of the stimulus utterances, it can be concluded that the time between the successive beats in the rhythm of an utterance can be measured. From this conclusion it follows that the hypothesis that conversational English is a stress-timed language may be tested.

Appendix A - Spectrograms and Mingograms

Pages 117 through 130 of Appendix A show spectrograms and mingograms of eight of the nine utterances used in the experiments described in Chapter II.* The spectrograms were made on a BTL Model D spectrograph and have been reduced by one-fourth. The mingograms show the speech wave, speech power, and a timing signal of 100 cps. The traces were made at 100 mm per sec and have been reduced by five-eighths. Because the timing pulse and the entire utterance could not both be included in a single spectrogram for some of the utterances, two spectrograms were made for these utterances. The lettering on the spectrograms and mingograms follows the orthographic spelling of the utterances and is intended only as a guide to the phonetic material. A close phonetic transcription was not made. The vertical line labeled "COMP LINE" on the mingograms indicates the point of comparison of the mingogram records used in Experiments 1 and 2 (see Section 3.4.3, p.70).

Pages 131 through 169 of Appendix A show locations in the speech of the means of the distributions of taps and click placements by all of the subjects. The locations are drawn on unreduced two inch sections of the spectrograms and mingograms. The syllables are numbered for reference from 1 to 97; this numbering is given on pages 115 and 116. The subjects in Experiment 1 are numbered 1, 2, and 3. Those in Experiment 2 are numbered I and II. Subject 1 was subject I and subject 3 was subject II. The line locating the mean of a subject's responses to a given syllable is labelled with the syllable number and the subject number, separated by a hyphen.

---

*The tape loop of utterance #3 used in Experiment 1 was lost.

Contents of Appendix A

Pages 117 through 130 show spectrograms and mingograms of utterances #1 through #9, excluding #3.

Pages 131 through 169 show locations of subjects' responses to all the syllables. Two syllables are included on each page, with the following syllable numbering·

Syllable	Number	Page	Syllable	Number	Page
use	1		in	21	140
the	2		cur	22	140
weight	3	131	rents	23	141
of	4	131	see	24	
the	5	132	when	25	
line	6	132	he	26	
to	7	133	reared	27	
get	8	133	back	28	
more$_1$	9	134	and	29	
and	10	134	fired	30	
more$_2$	11	135	it	31	
out	12	135	by	32	
spin	13	136	a	33	
ners	14	136	guy	34	
are	15	137	af	35	
par	16	137	ter	36	
tic	17	138	the	37	141
u	18	138	play	38	142
ly	19	139	was	39	142
good	20	139	o	40	143

Syllable	Number	Page	Syllable	Number	Page	Syllable	Number	Page
ver	41	143	I	63	153	what	85	
he	42	144	would	64	154	will	86	
wants	43	144	like	65	154	hap	87	164
to	44	145	to	66	155	pen	88	164
be	45	145	know	67	155	now	89	165
a	46	146	how	68	156	and	90	165
per	47	146	they	69	156	sort	91	166
for	48	147	con	70	157	of	92	166
mer	49	147	struct	71	157	go	93	167
now	50	148	these	72	158	out	94	167
talks	51		things	73	158	on	95	168
in	52		I_1	74		a	96	168
terms	53	148	like	75	159	limb	97	169
of	54	149	pre_1	76	159			
get	55	149	dict_1	77	160			
ting	56	150	ing	78	160			
a	57	150	I_2	79	161			
name	58	151	love	80	161			
for	59	151	to	81	162			
him	60	152	pre_2	82	162			
self	61	152	dict_2	83	163			
so	62	153	things	84	163			

Fig. A1. Spectrogram of Utterance 1 Showing Timing Pulse

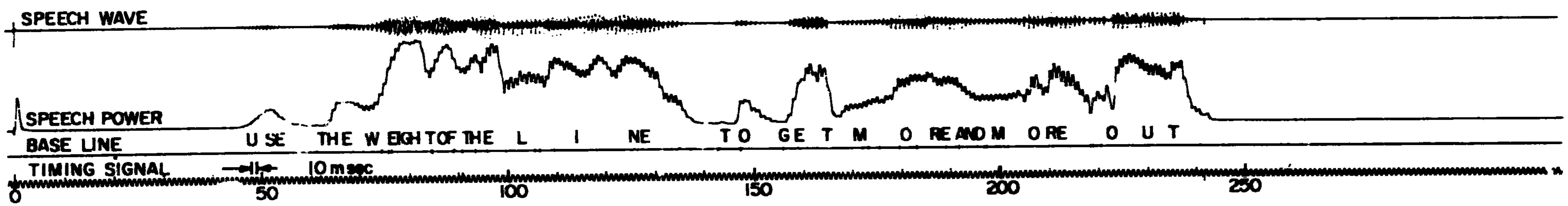

Fig. A2. Mingogram of Utterance 1

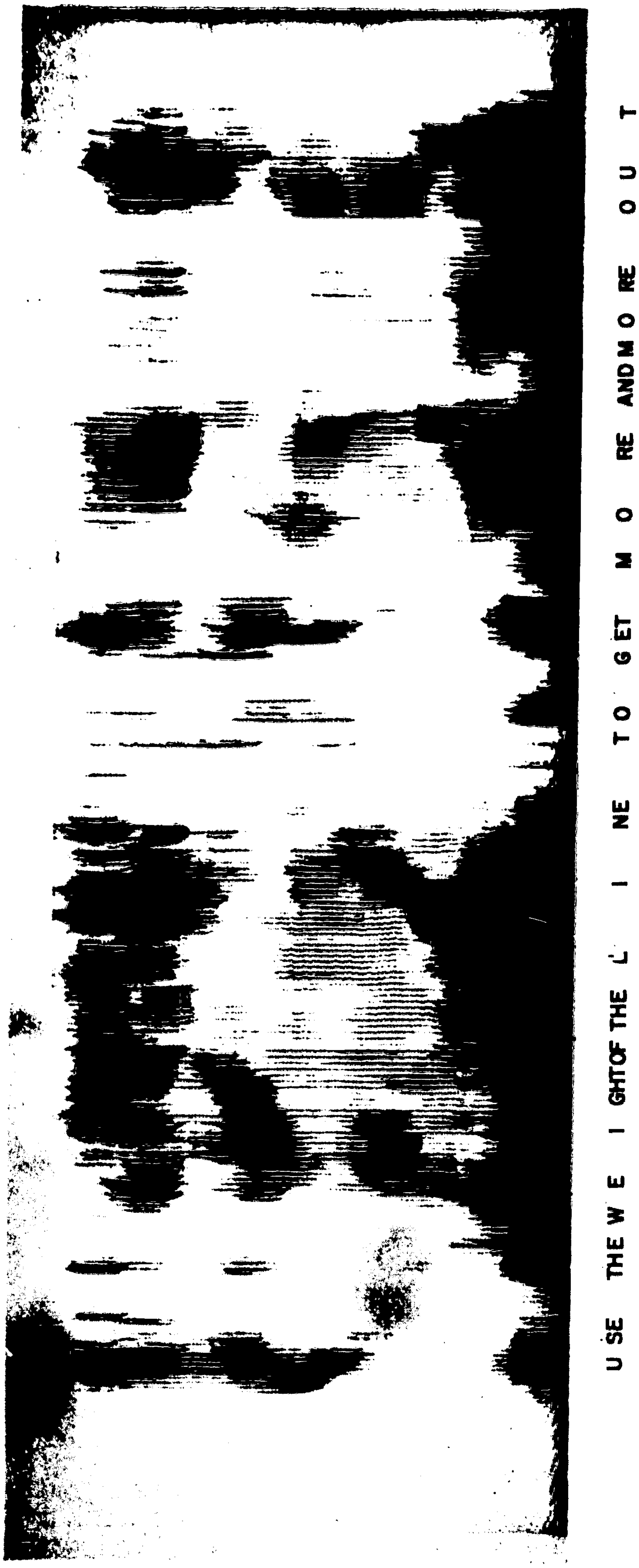

Fig. A3. Spectrogram of Utterance 1

Fig. A4. Spectrogram of Utterance 2 Showing Timing Pulse

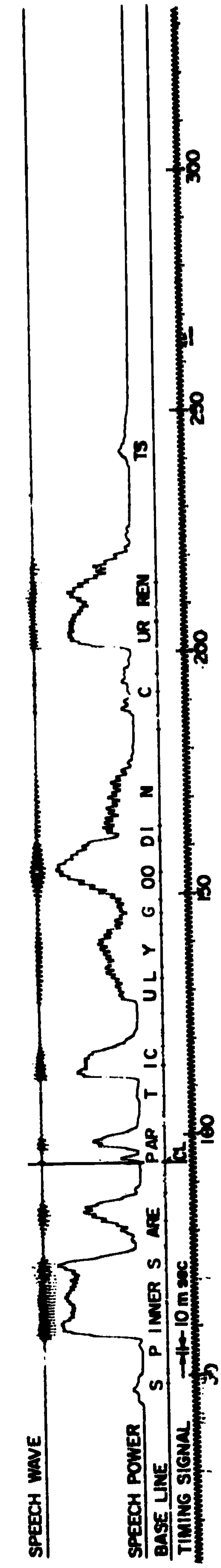

Fig. A5. Mingogram of Utterance 2

Fig. A6. Spectrogram of Utterance 2

Fig. A7. Spectrogram of Utterance 4 Showing Timing Pulse

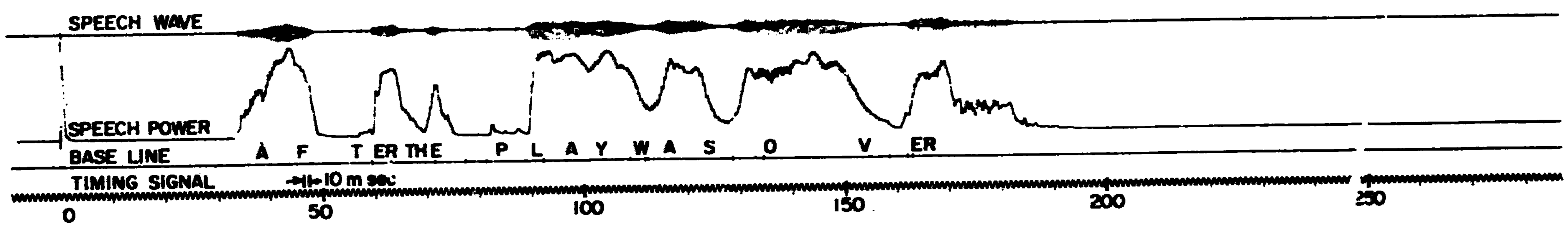

Fig. A8. Mingogram of Utterance 4

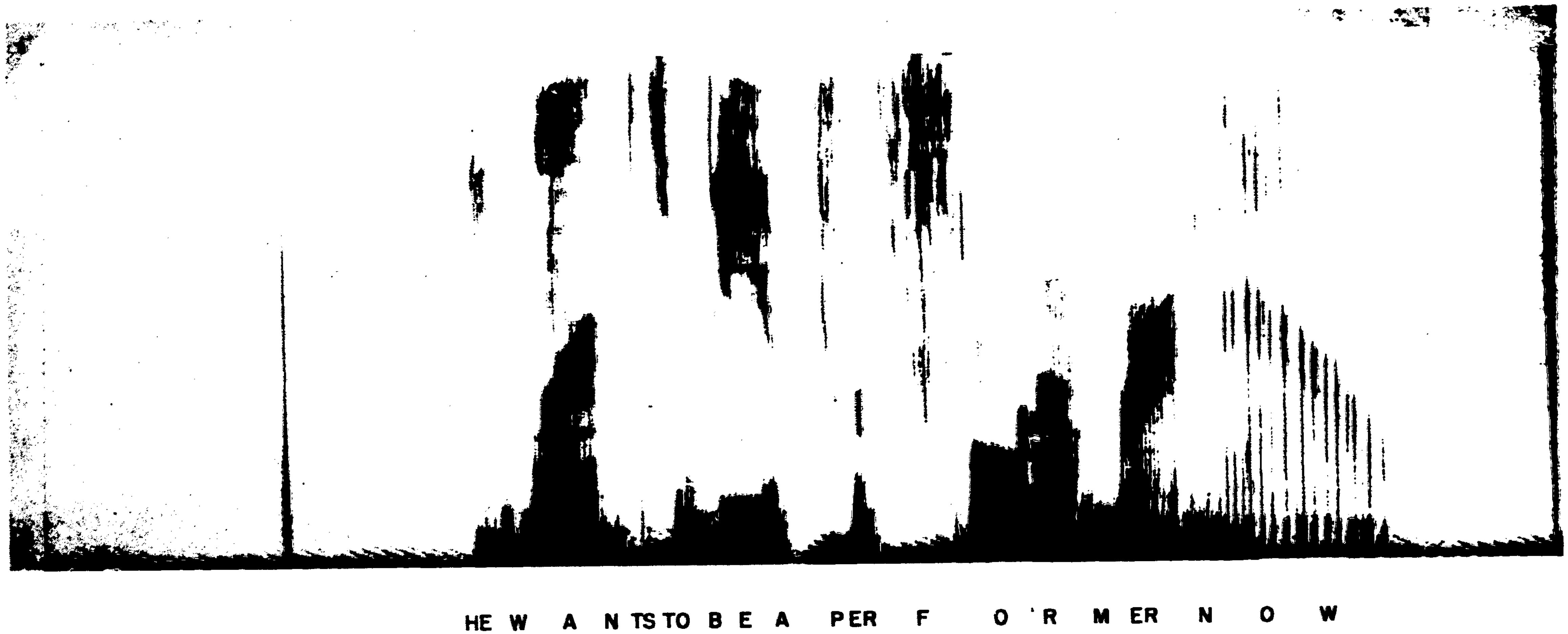

Fig. A9. Spectrogram of Utterance 5 Showing Timing Pulse

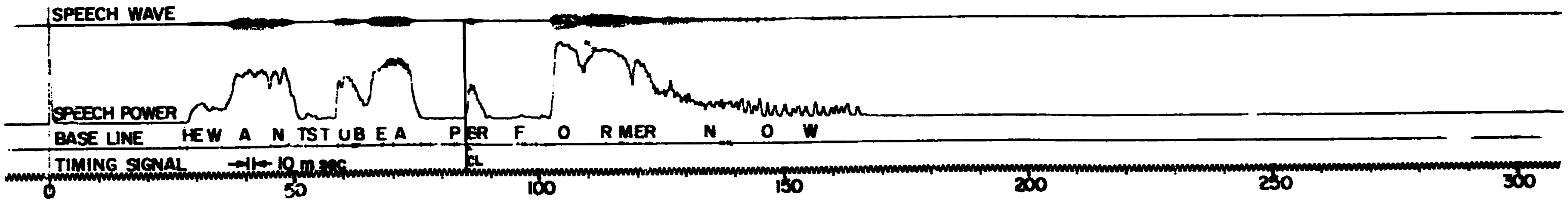

Fig. A10. Mingogram of Utterance 5

Fig. A11. Spectrogram of Utterance 6 Showing Timing Pulse

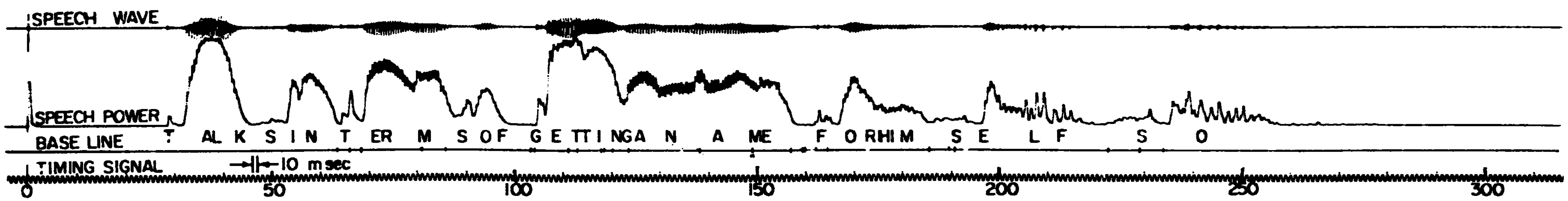

Fig. A12. Mingrogram of Utterance 6

G ETT I NGA N A ME F O R HI M S E L F S O

Fig. A13. Spectrogram of Utterance 6

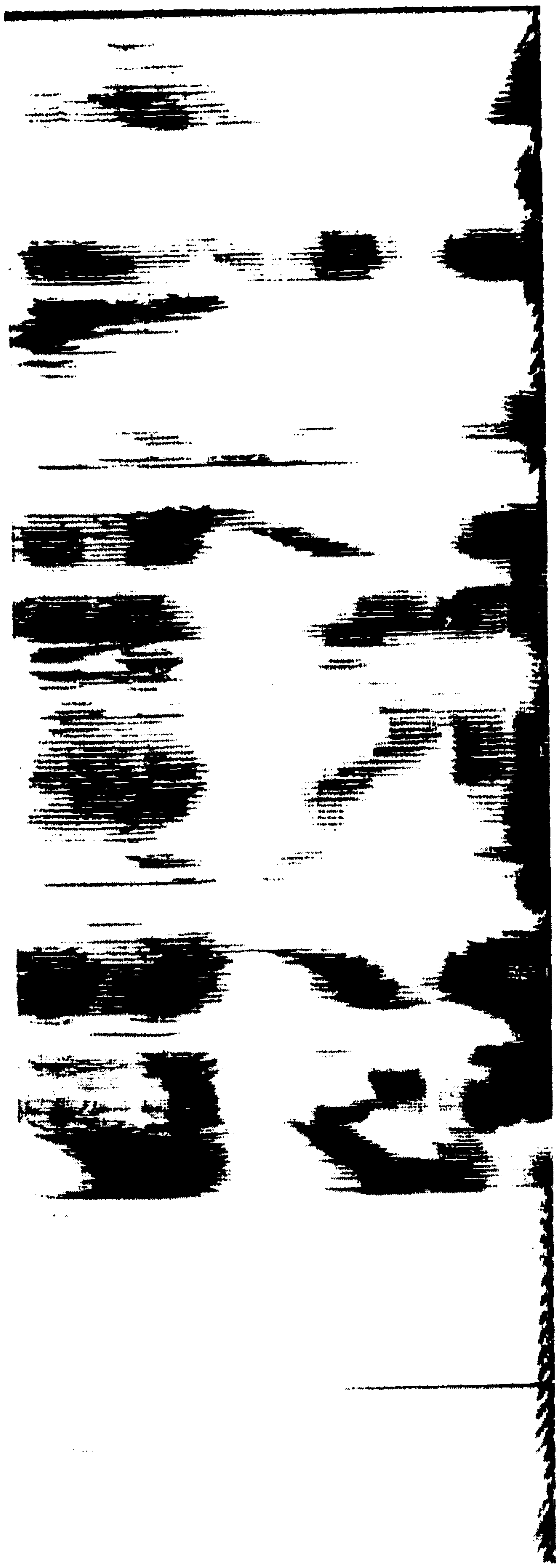

Fig. A14. Spectrogram of Utterance 7 Showing Timing Pulse

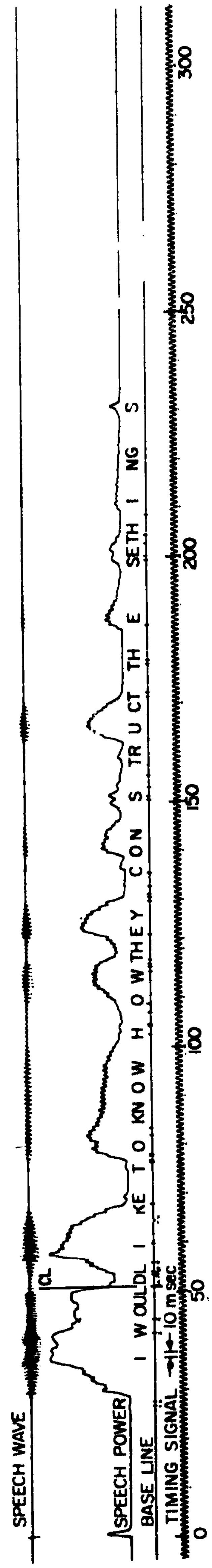

Fig. A15. Mingogram of Utterance 7

Fig. A16. Spectrogram of Utterance 7

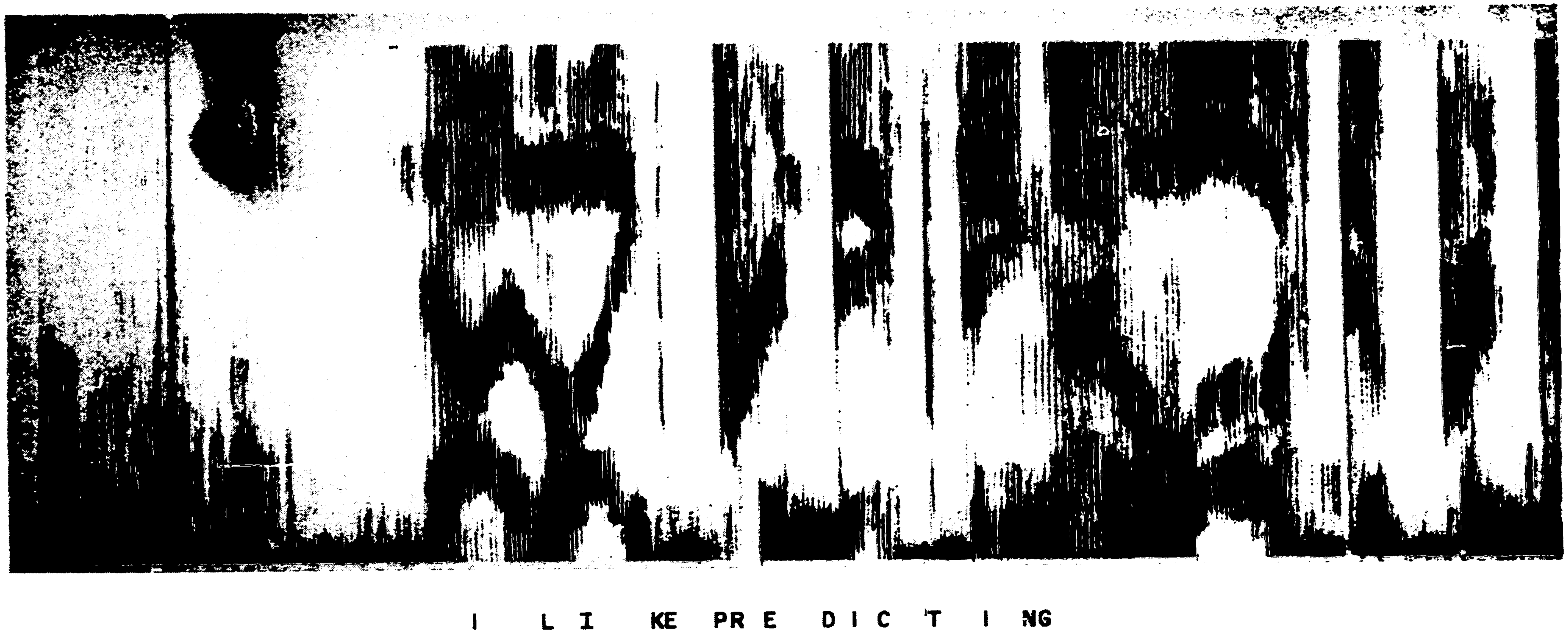

Fig. A17. Spectrogram of Utterance 8 Showing Timing Pulse

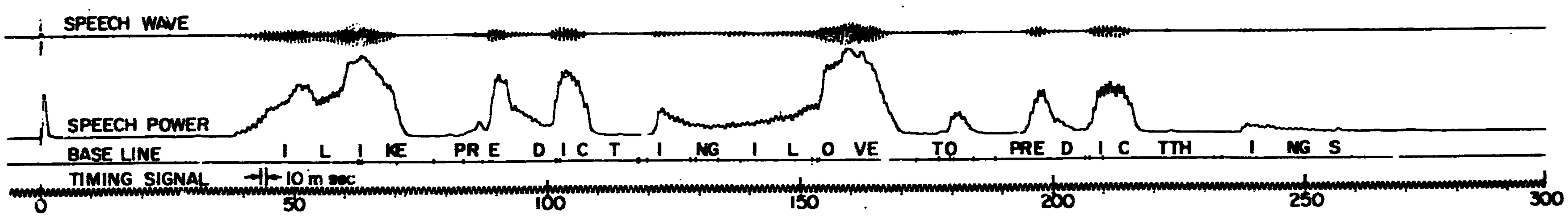

Fig. A18. Mingogram of Utterance 8

I · L O VE TO PRE D I C T TH I NG S

Fig. A19. Spectrogram of Utterance 8

WH A T LL H A PP E N N O W

Fig. A20. Spectrogram of Utterance 9 Showing Timing Pulse

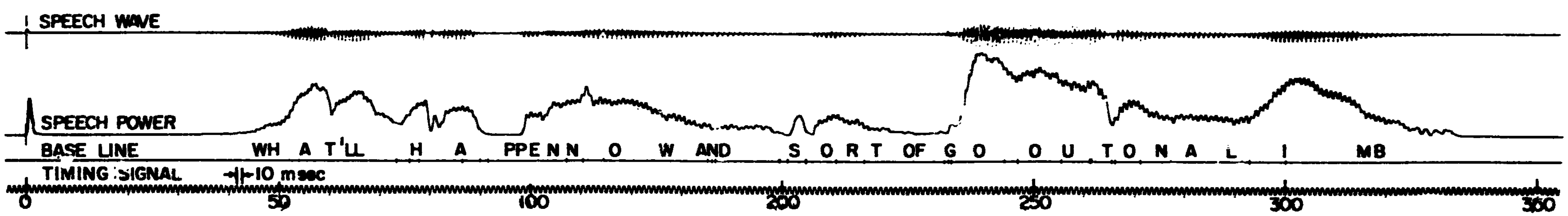

Fig. A21. Mingogram of Utterance 9

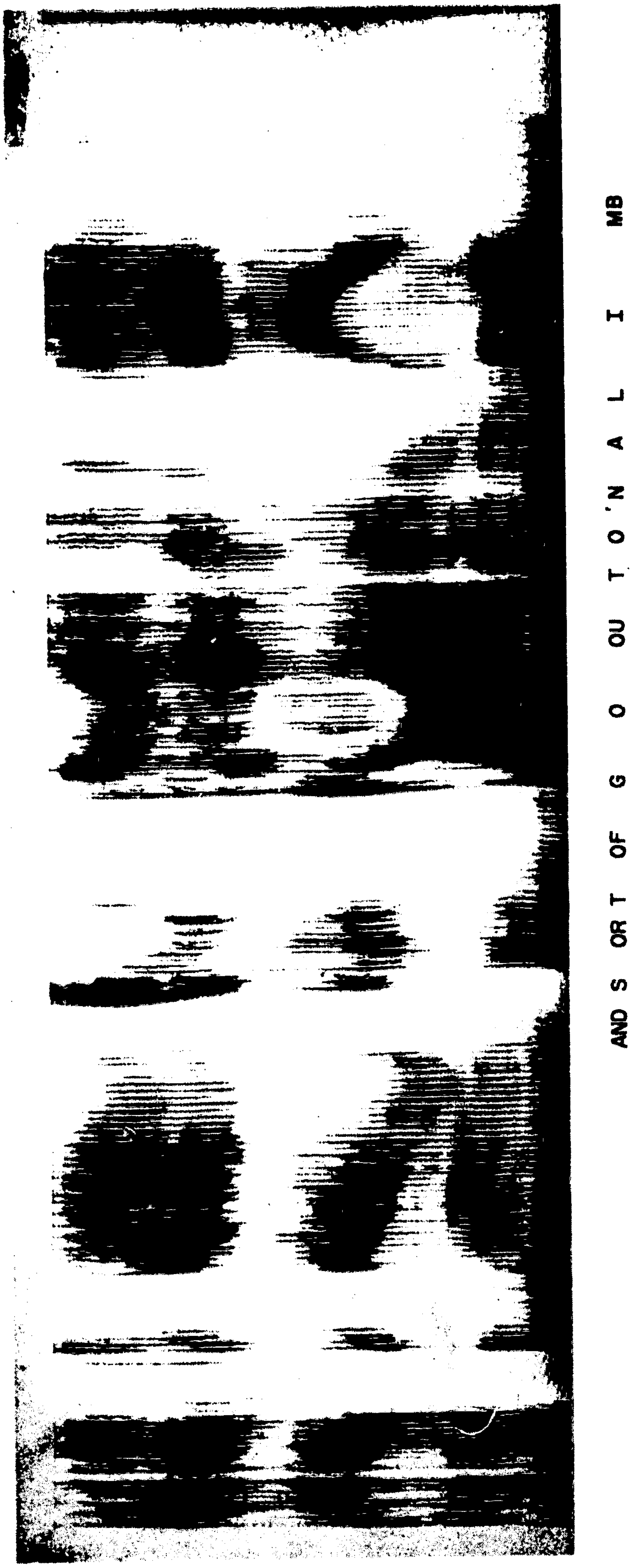

Fig. A22. Spectrogram of Utterance 9

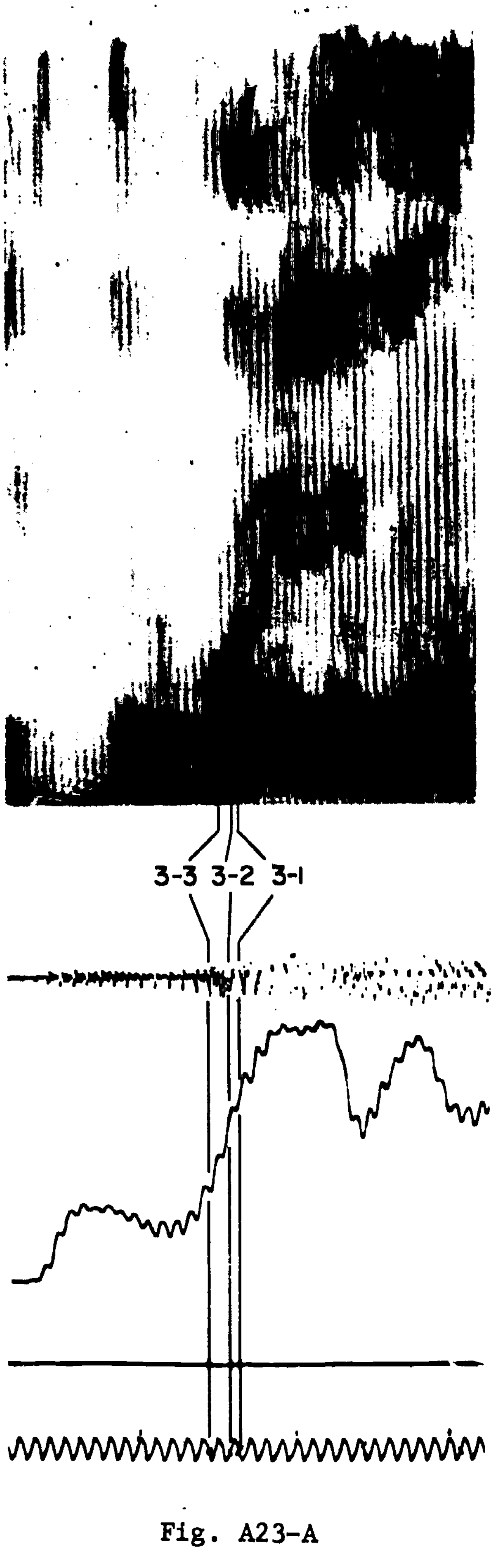

Fig. A23-A

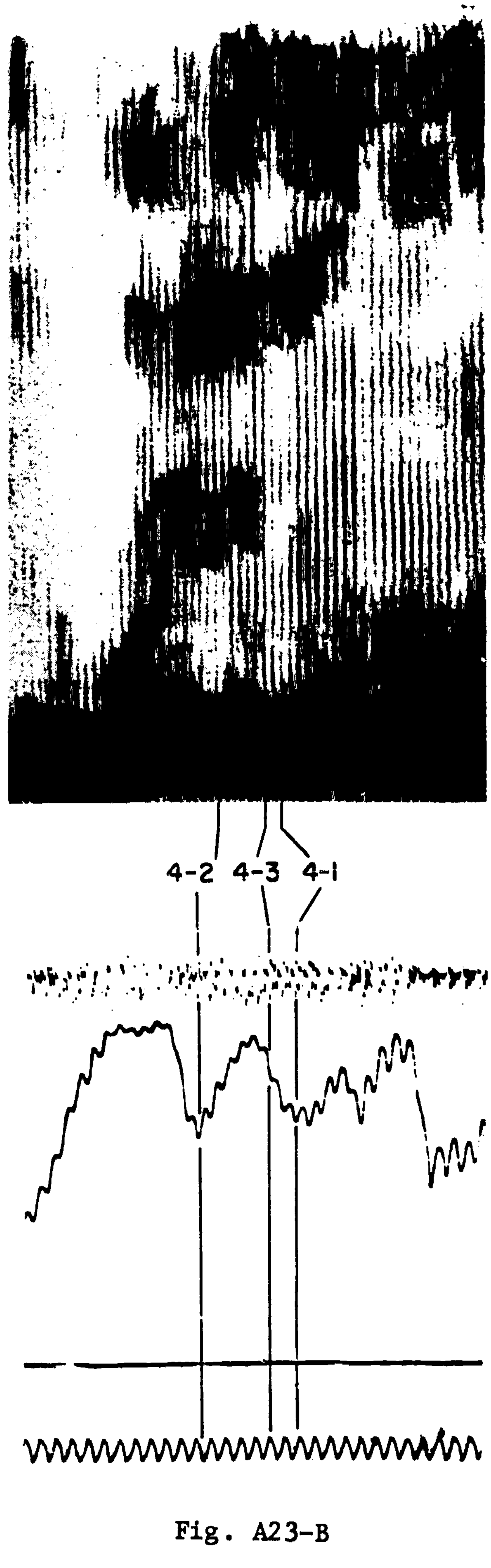

Fig. A23-B

Fig. A23-A,B: Spectrogram and Mingogram Sections
for Syllables 3 and 4

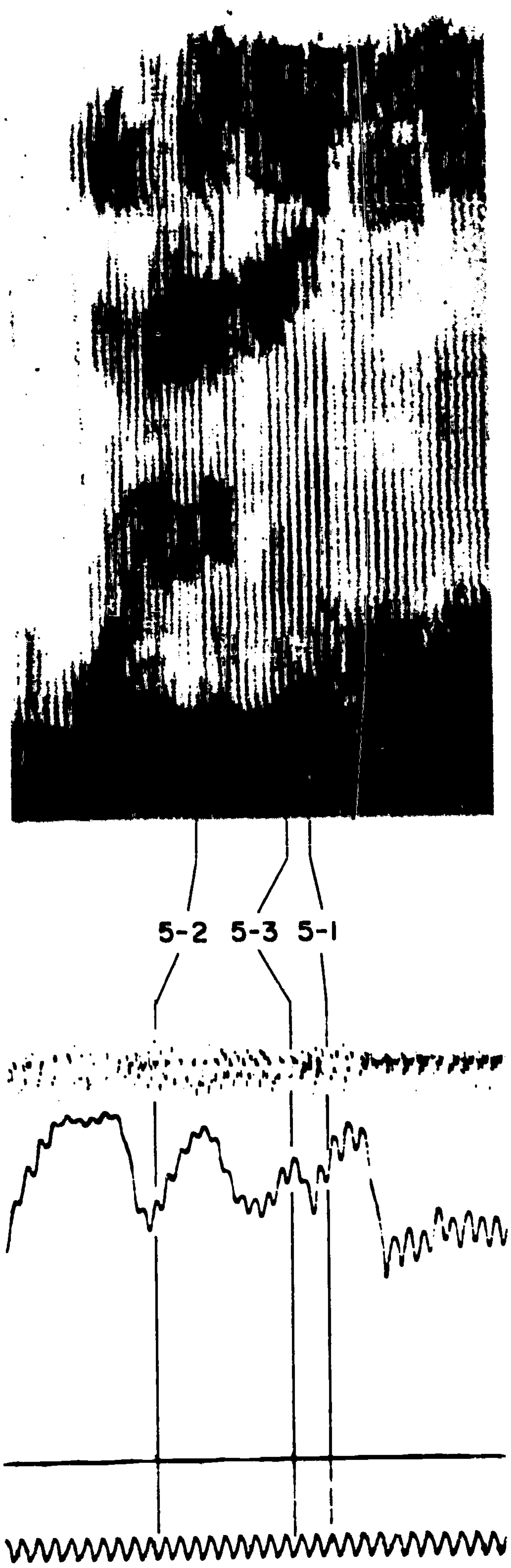

Fig. A24-A

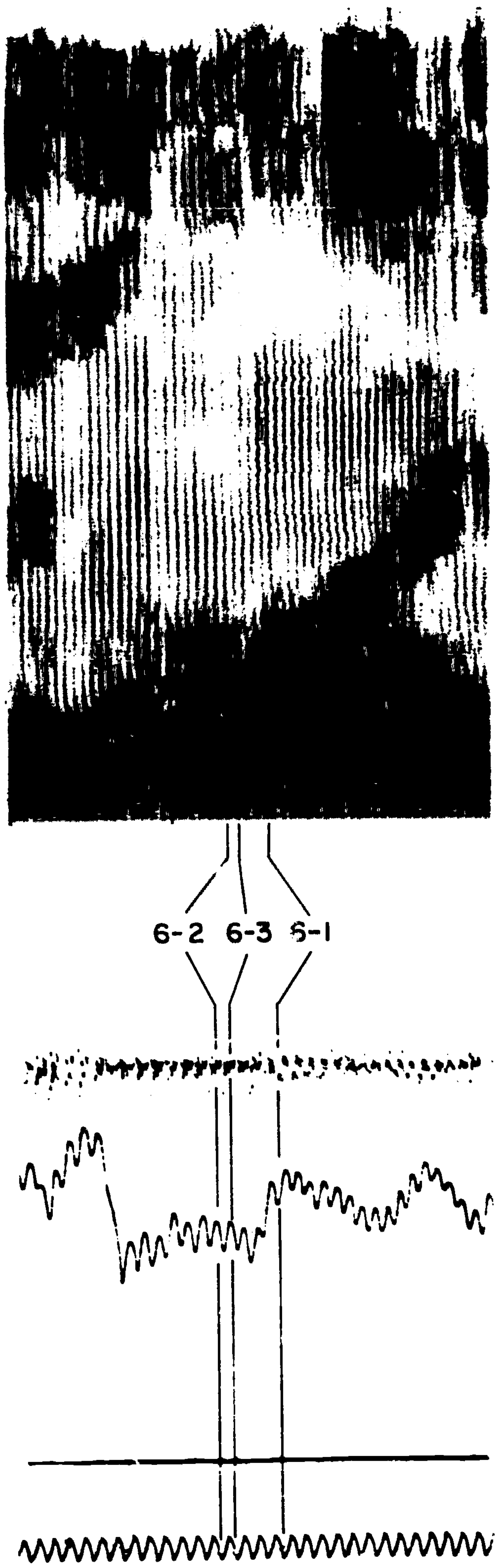

Fig. A24-B

Fig. A24-A,B: Spectrogram and Mingogram Sections for Syllables 5 and 6

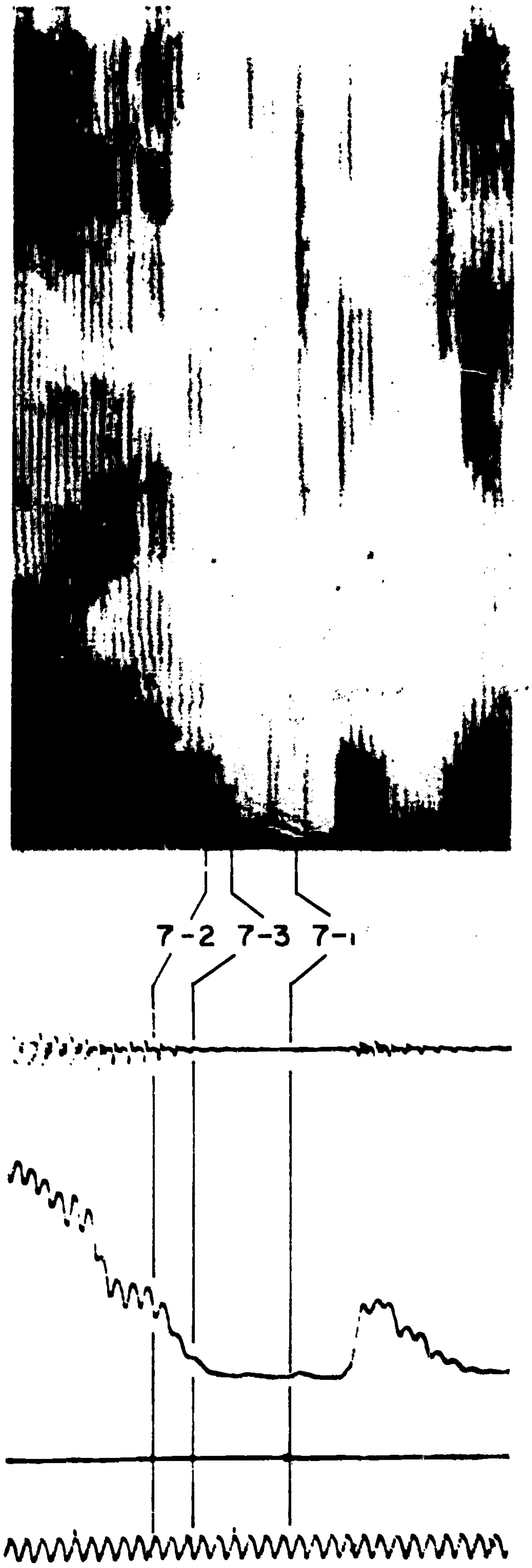

Fig. A25-A

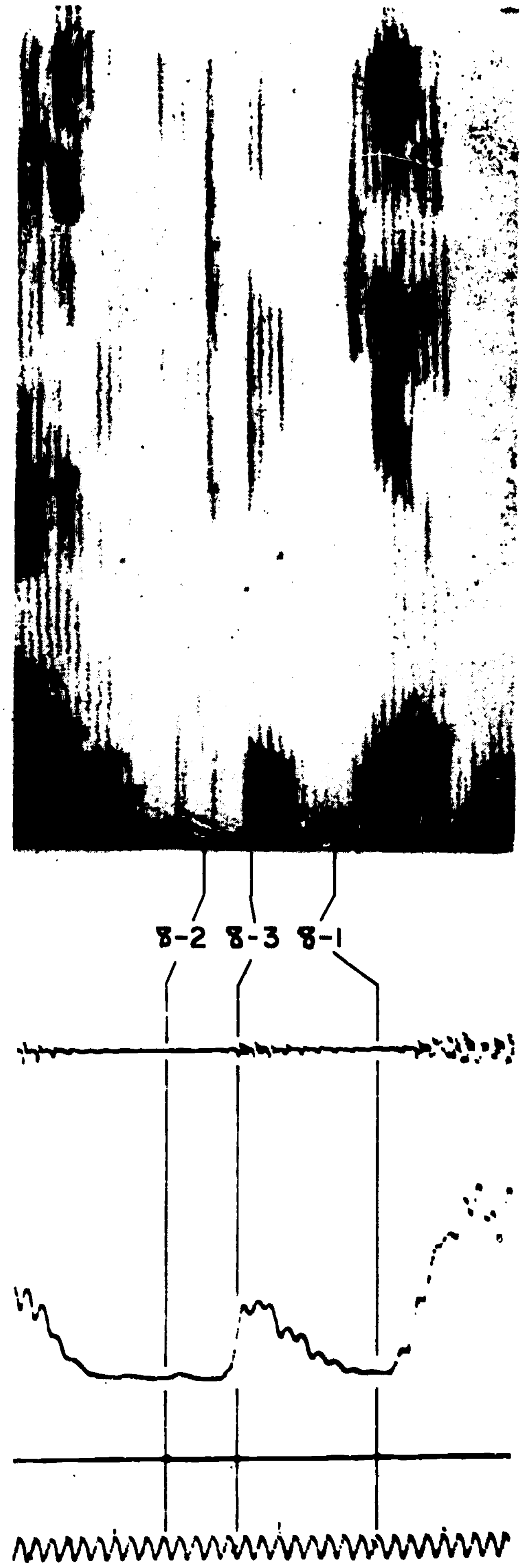

Fig. A25-B

Fig. A25-A,B: Spectrogram and Mingogram Sections for Syllables 7 and 8

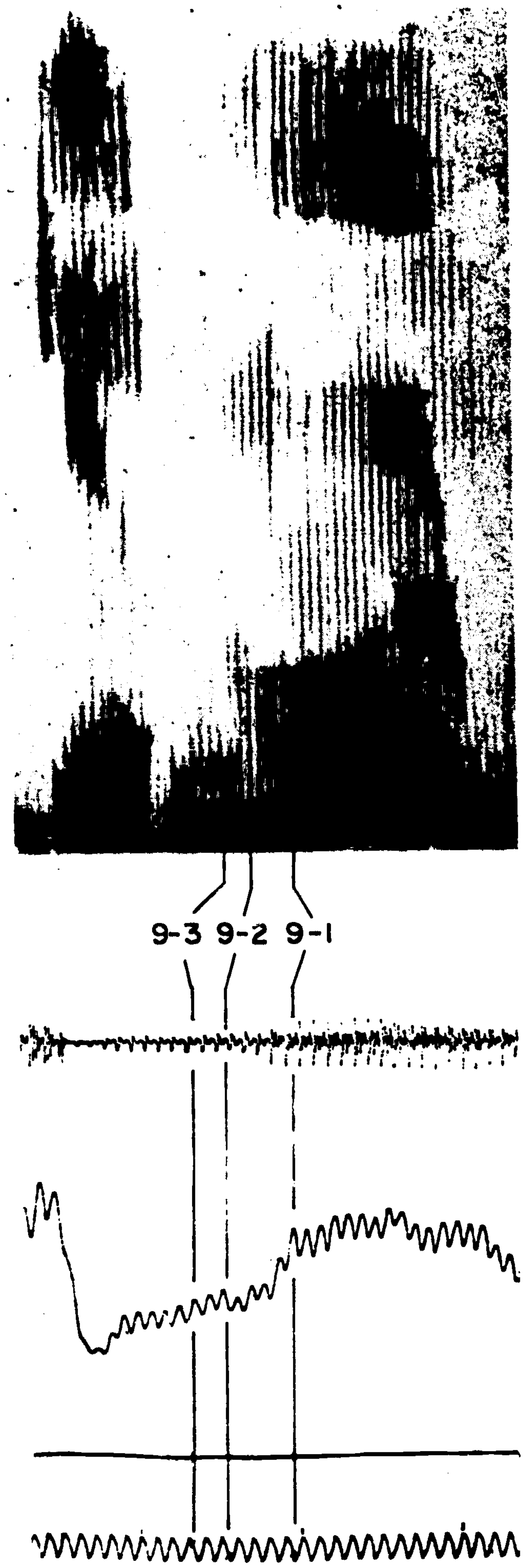

Fig. A26-A

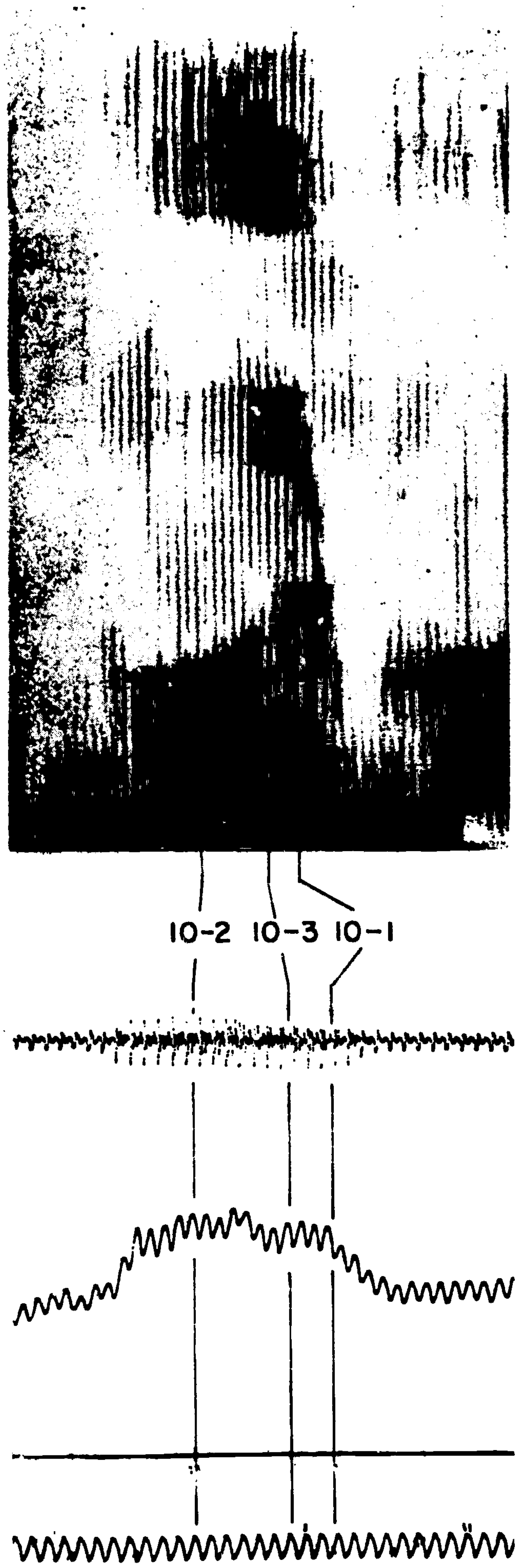

Fig. A26-B

Fig. A26-A,B: Spectrogram and Mingogram Sections for Syllables 9 and 10

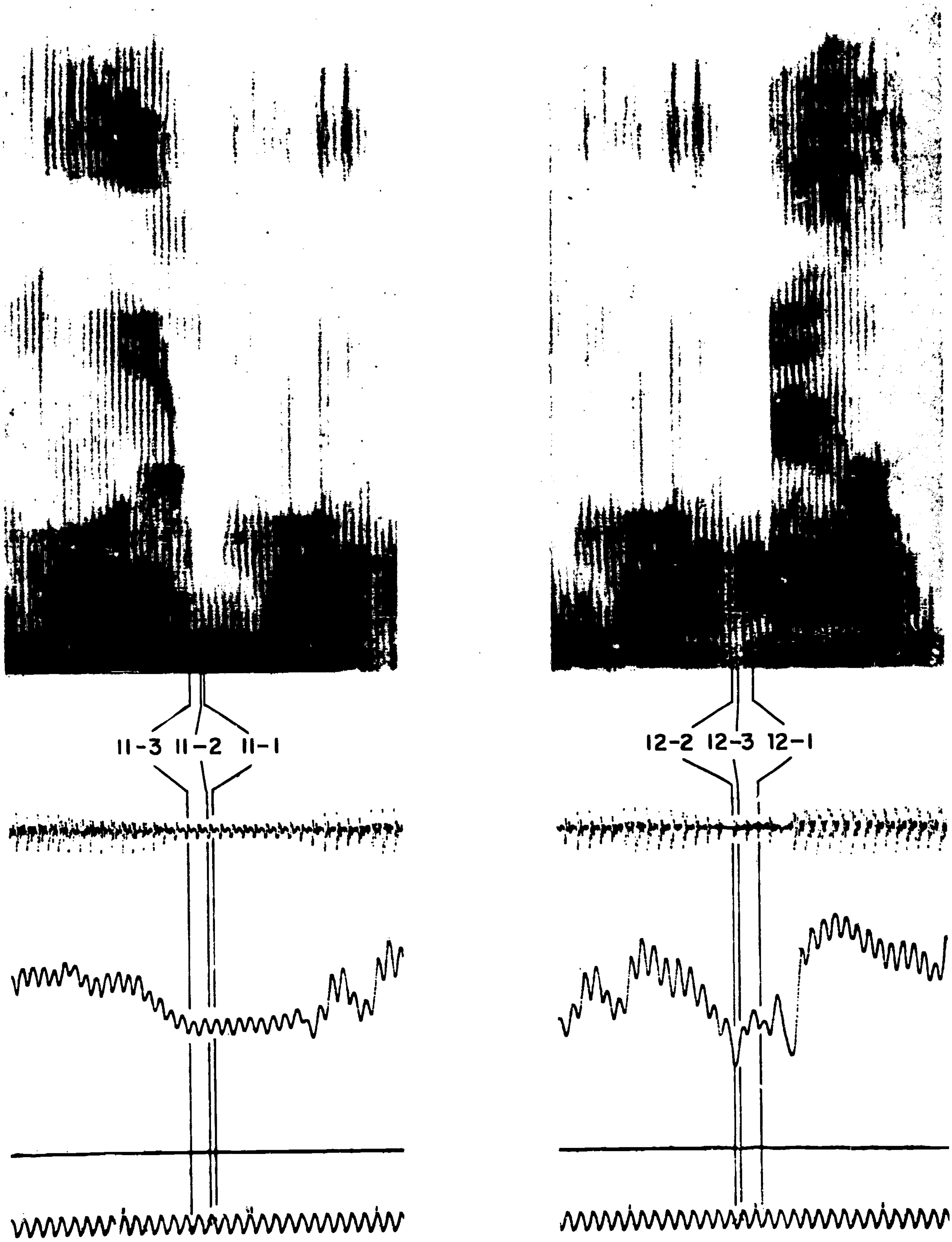

Fig. A27-A

Fig. A27-B

Fig. A27-A,B: Spectrogram and Mingogram Sections for Syllables 11 and 12

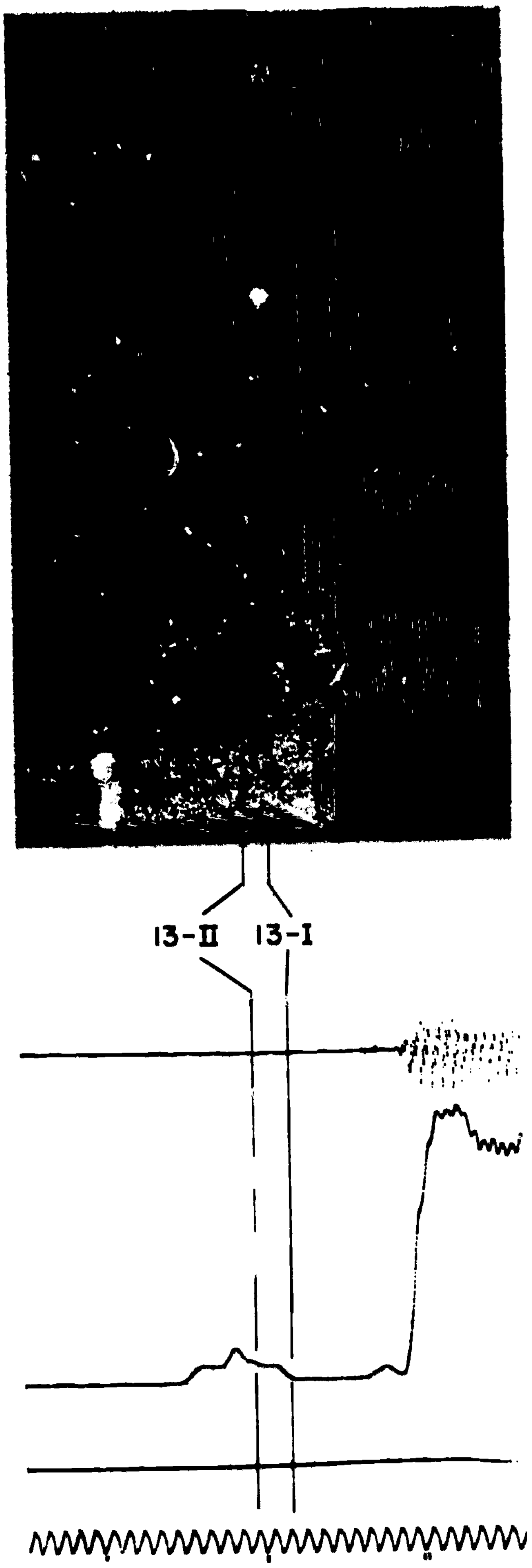

Fig. A28-A

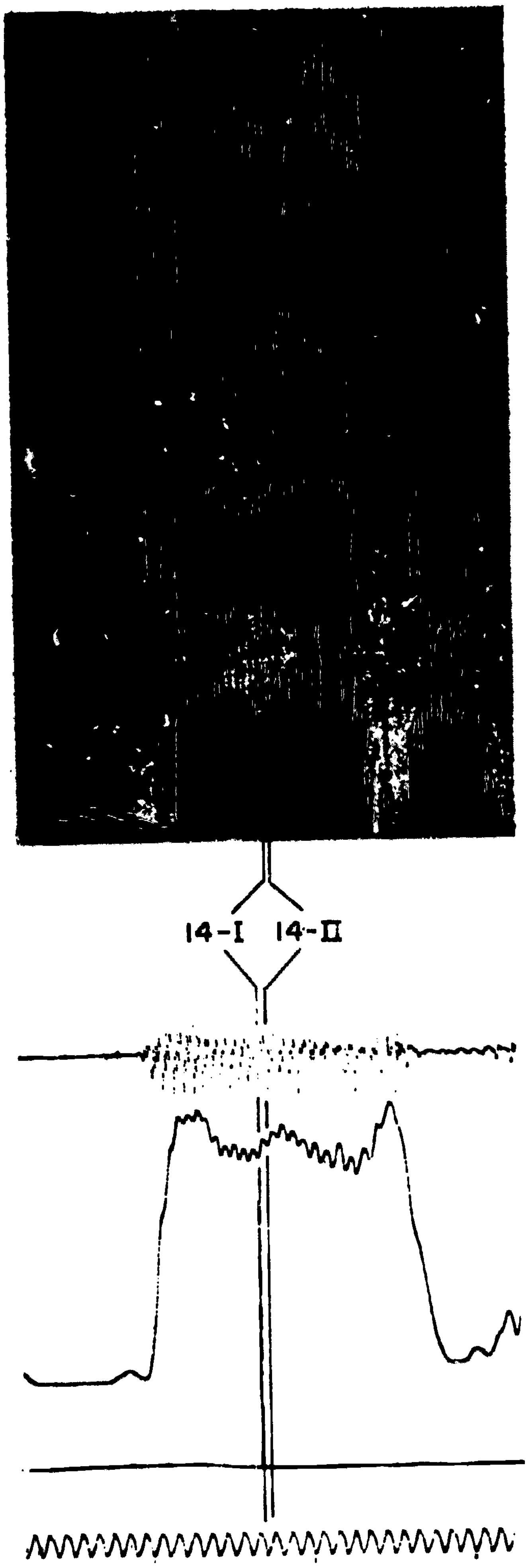

Fig. A28-B

Fig. A28-A,B: Spectrogram and Mingogram Sections for Syllables 13 and 14

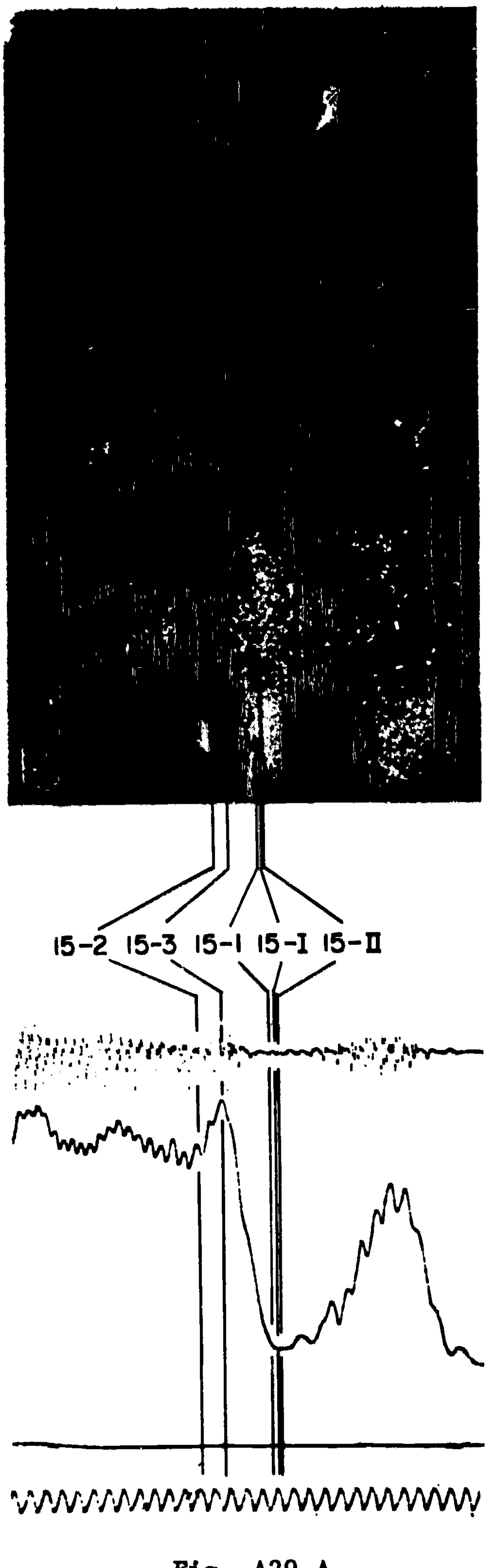

Fig. A29-A

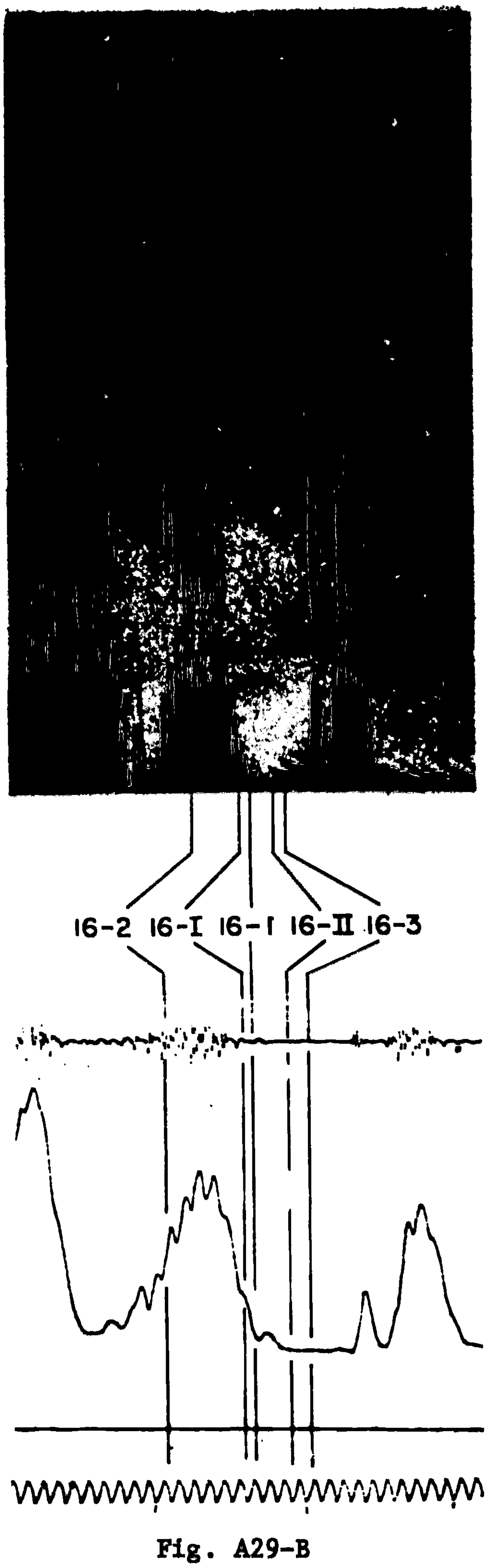

Fig. A29-B

Fig. A29-A,B: Spectrogram and Mingogram Sections for Syllables 15 and 16

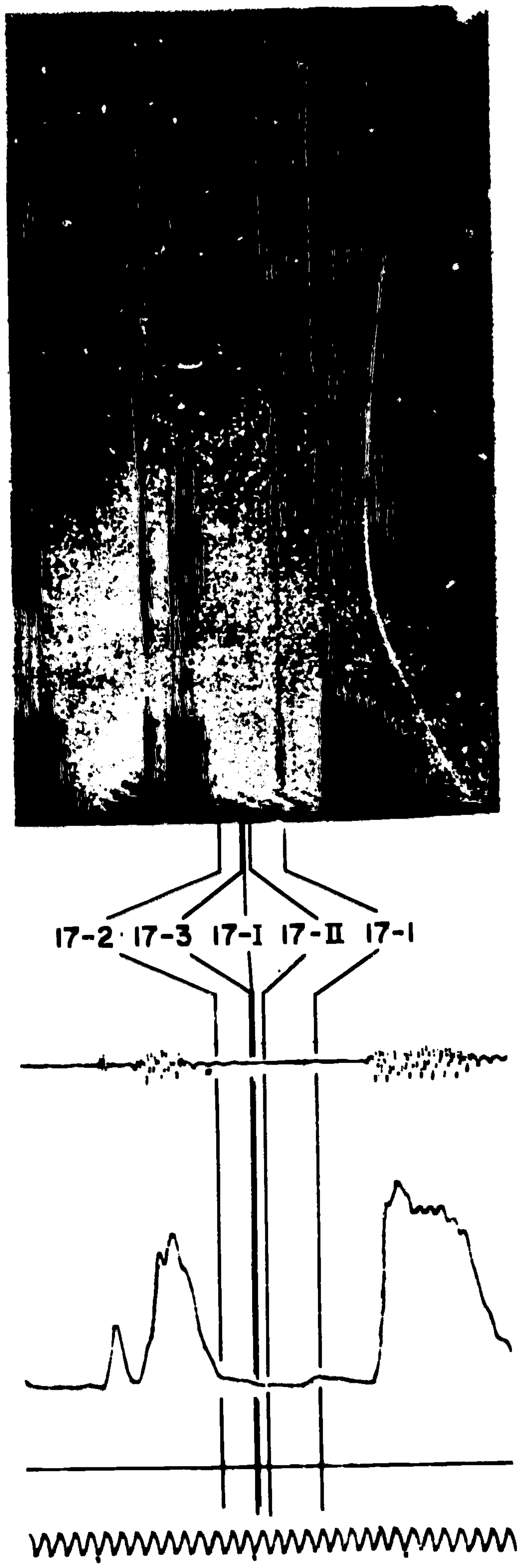

Fig. A30-A

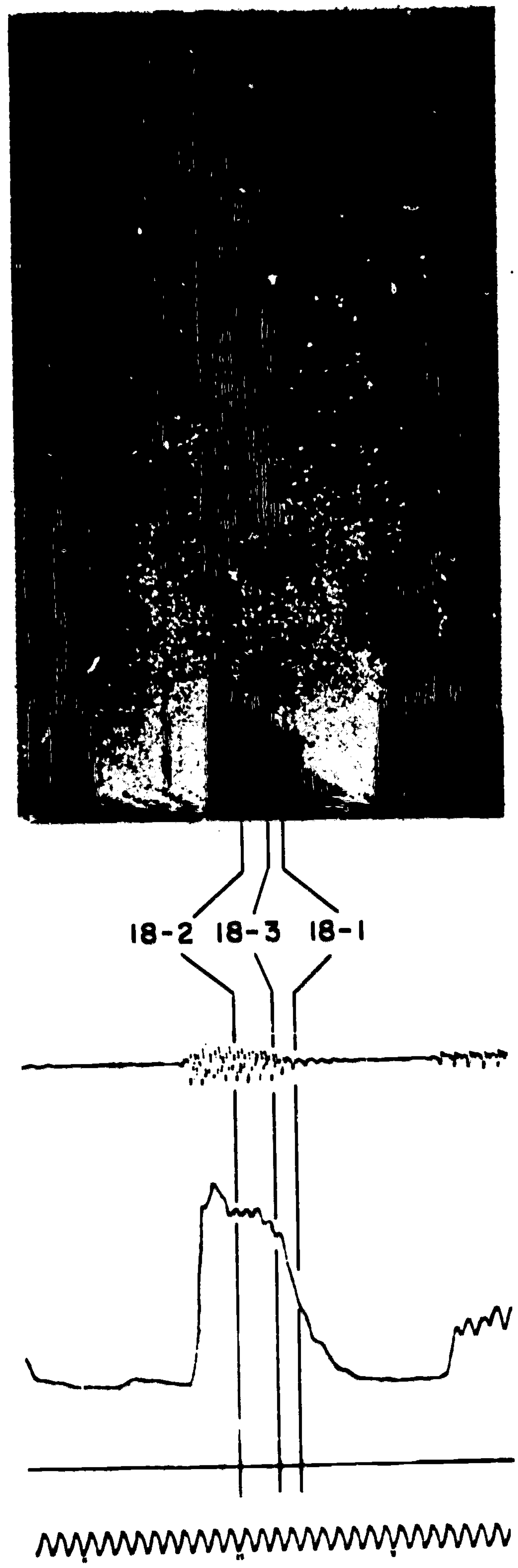

Fig. A30-B

Fig. A30-A,B: Spectrogram and Mingogram Sections for Syllables 17 and 18

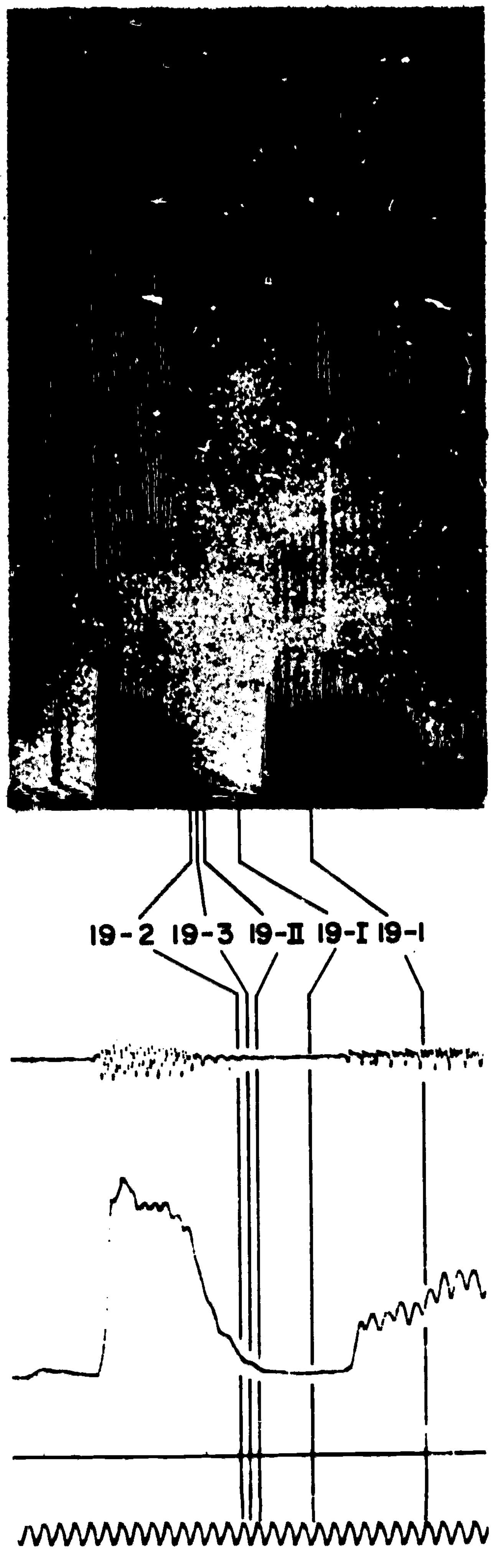

Fig. A31-A

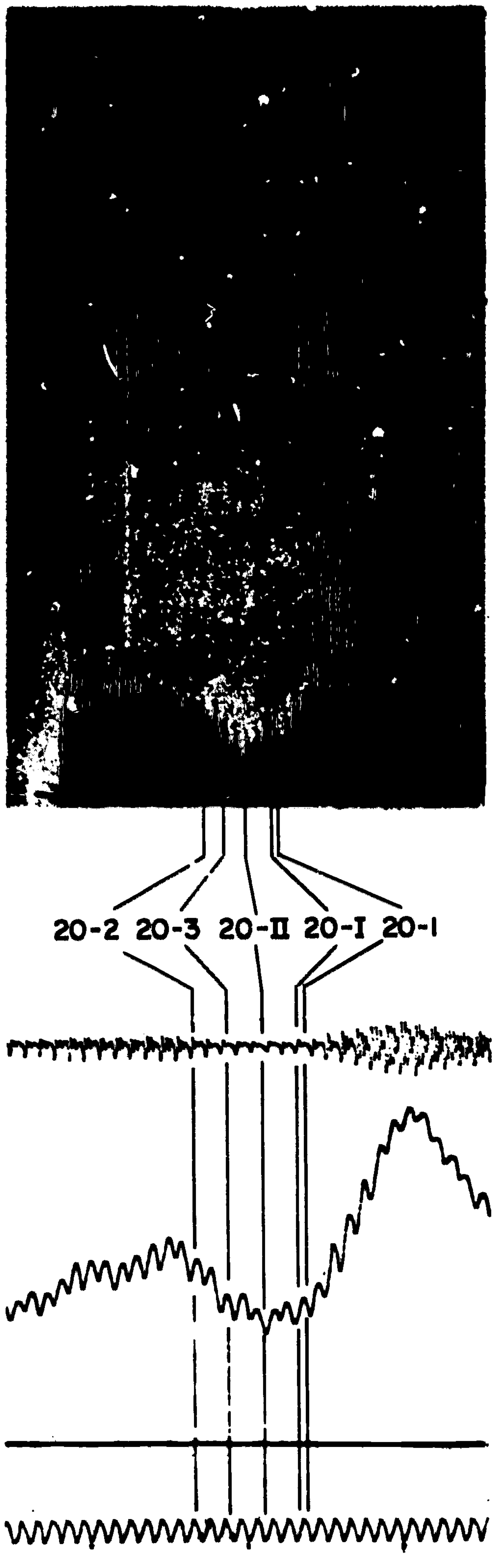

Fig. A31-B

Fig. A31-A,B: Spectrogram and Mingogram Sections for Syllables 19 and 20

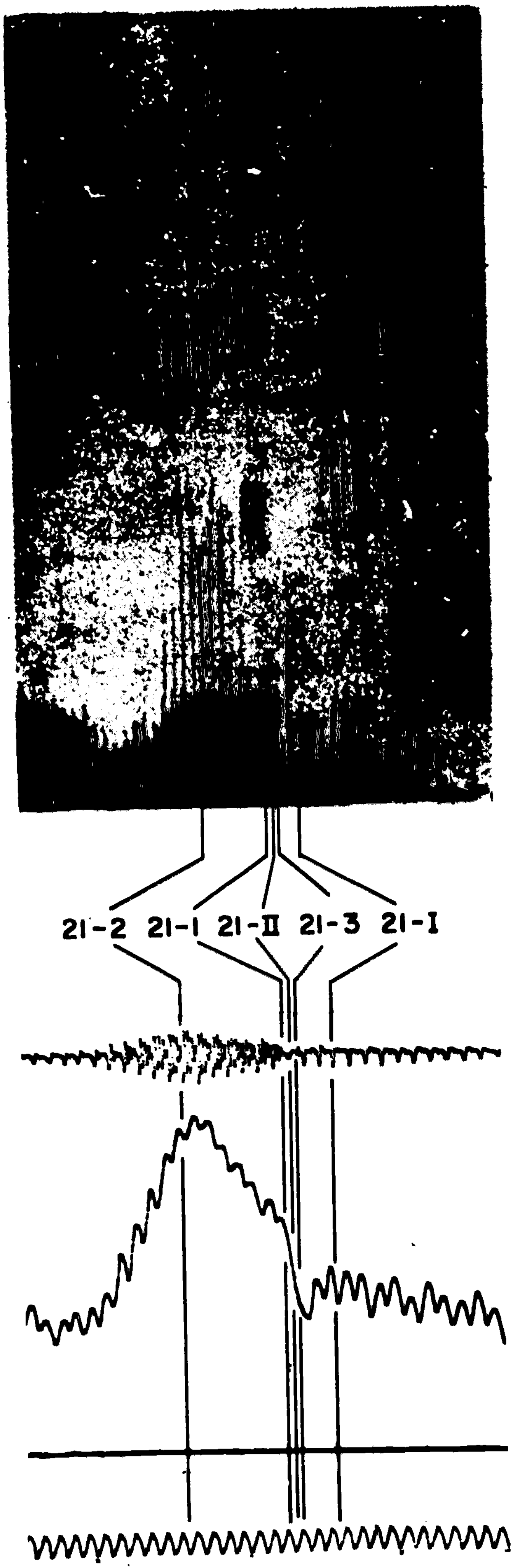

Fig. A32-A

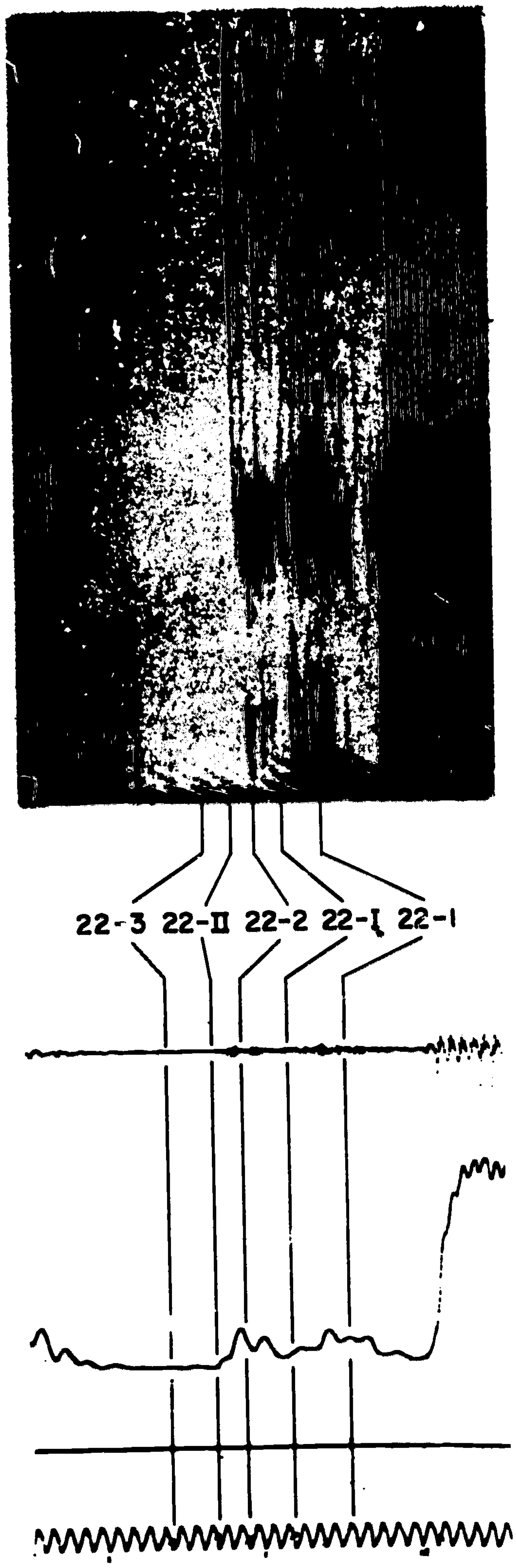

Fig. A32-B

Fig. A32-A,B: Spectrogram and Mingogram Sections for Syllables 21 and 22

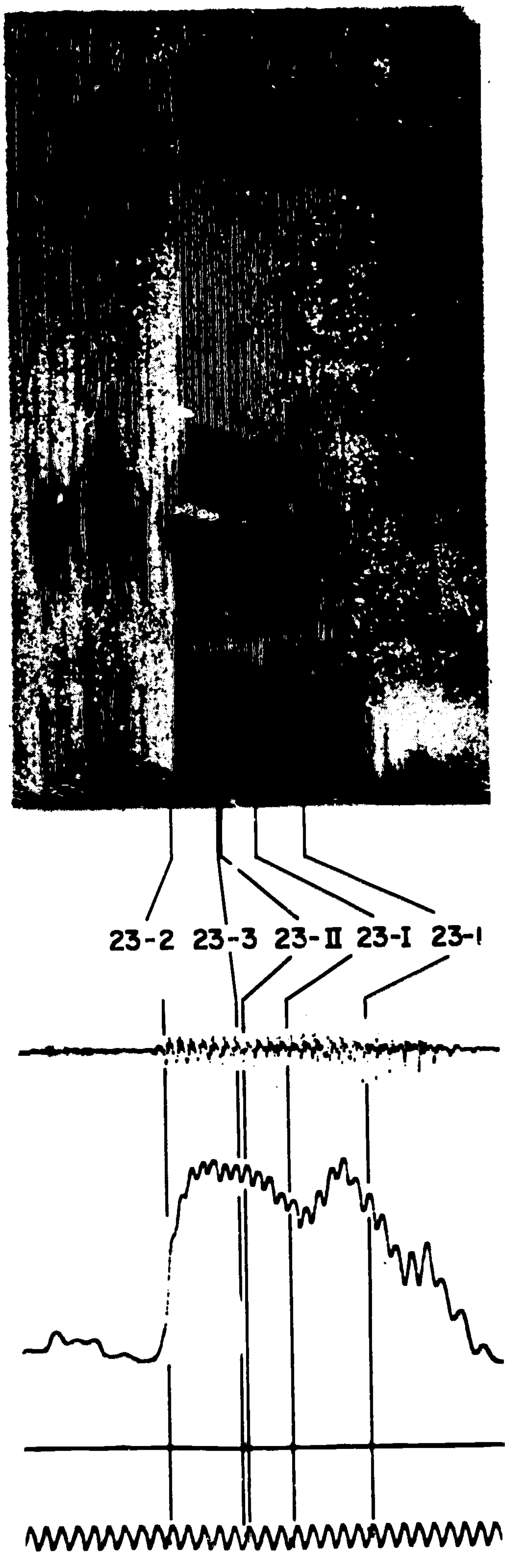

Fig. A33-A

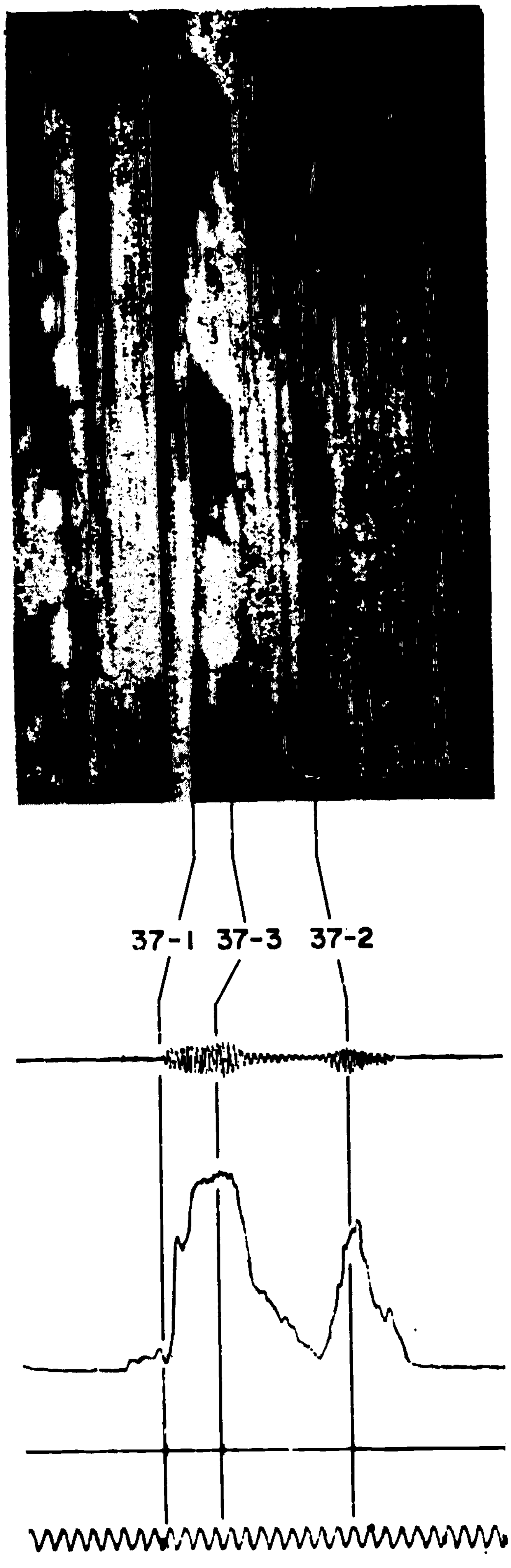

Fig. A33-B

Fig. A33-A,B: Spectrogram and Mingogram Sections for Syllables 23 and 37

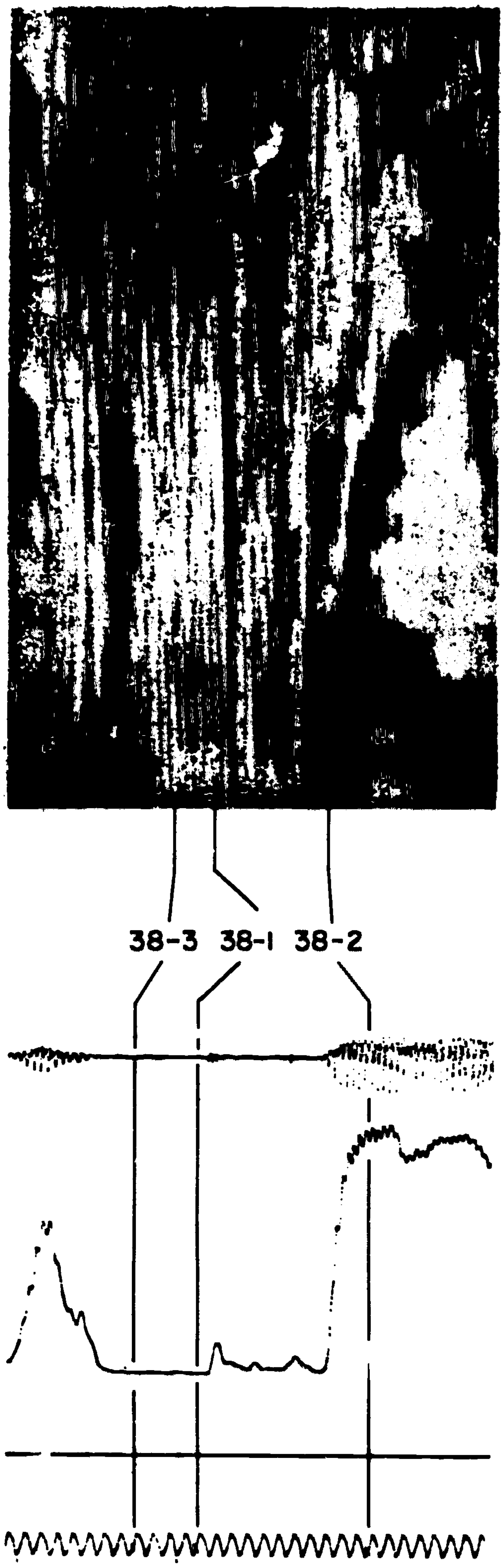

Fig. A34-A

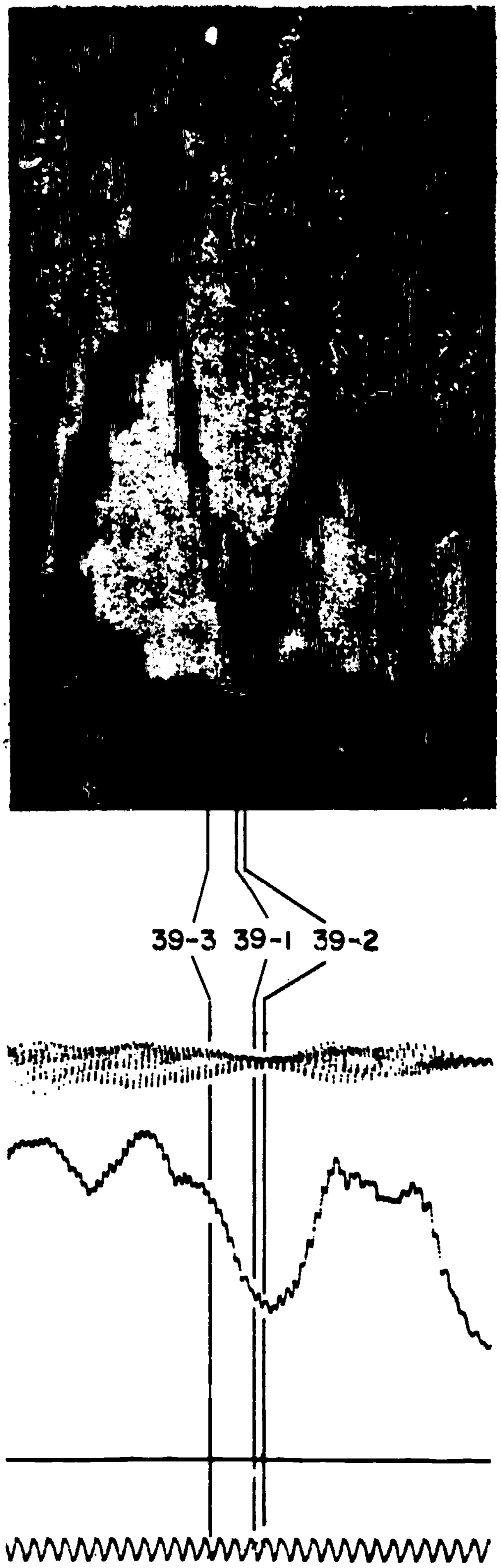

Fig. A34-B

Fig. A34-A,B: Spectrogram and Mingogram Sections
for Syllables 38 and 39

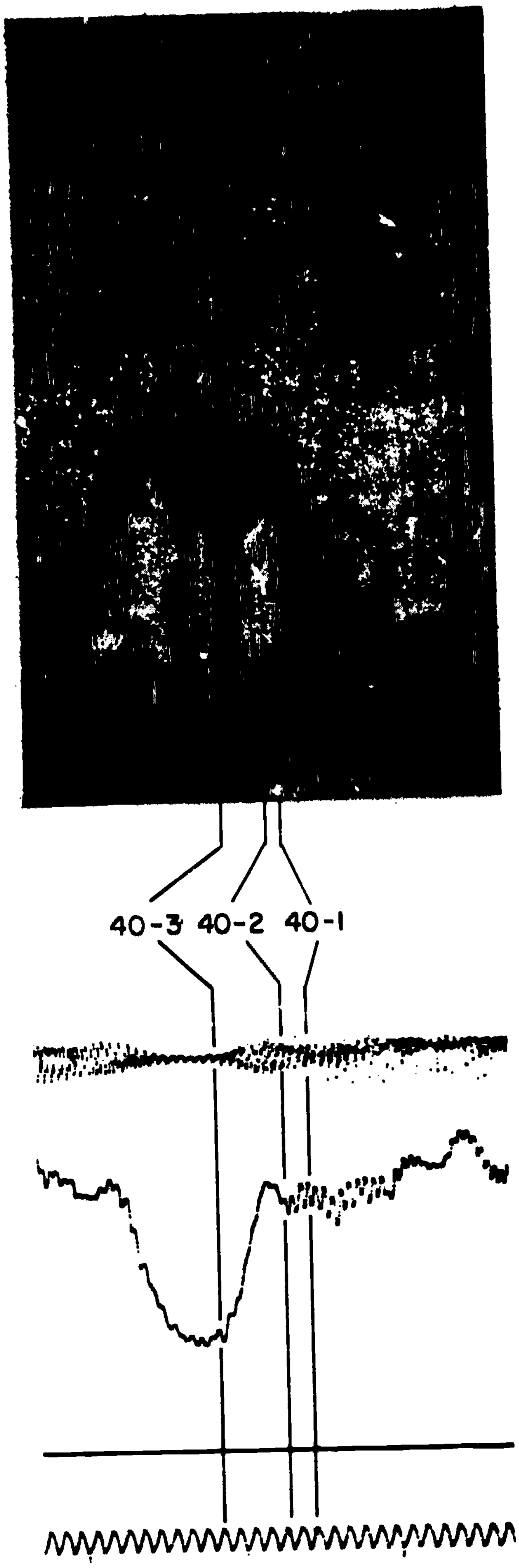

Fig. A35-A

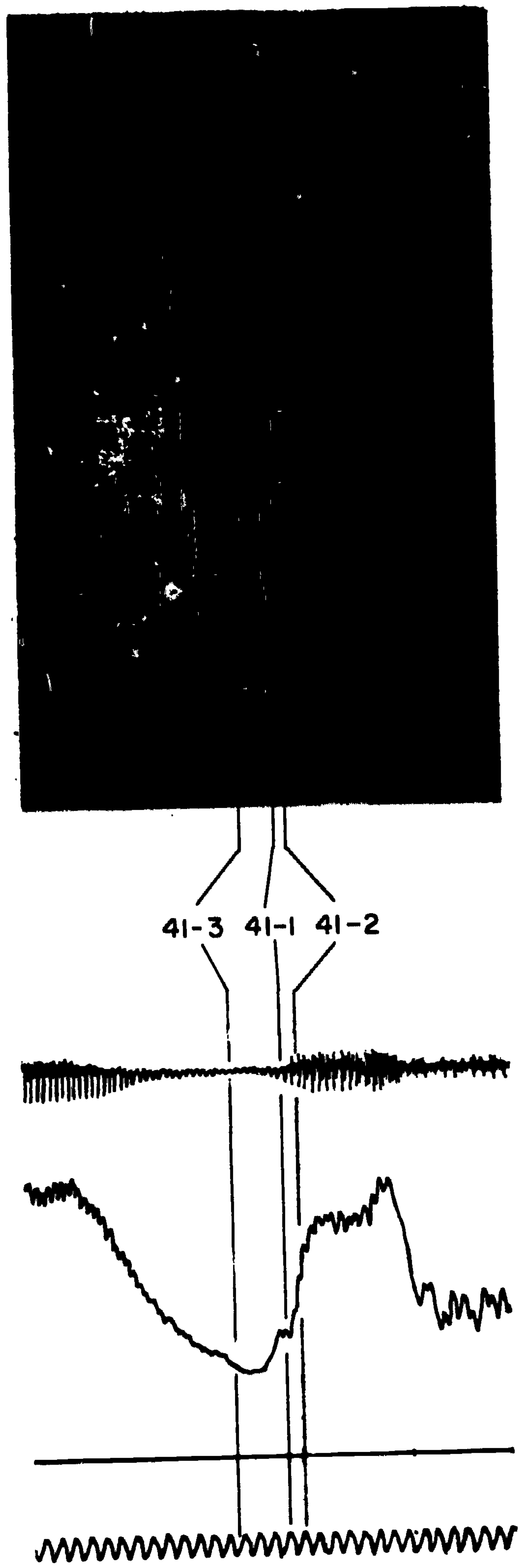

Fig. A35-B

Fig. A35-A,B: Spectrogram and Mingogram Sections for Syllables 40 and 41

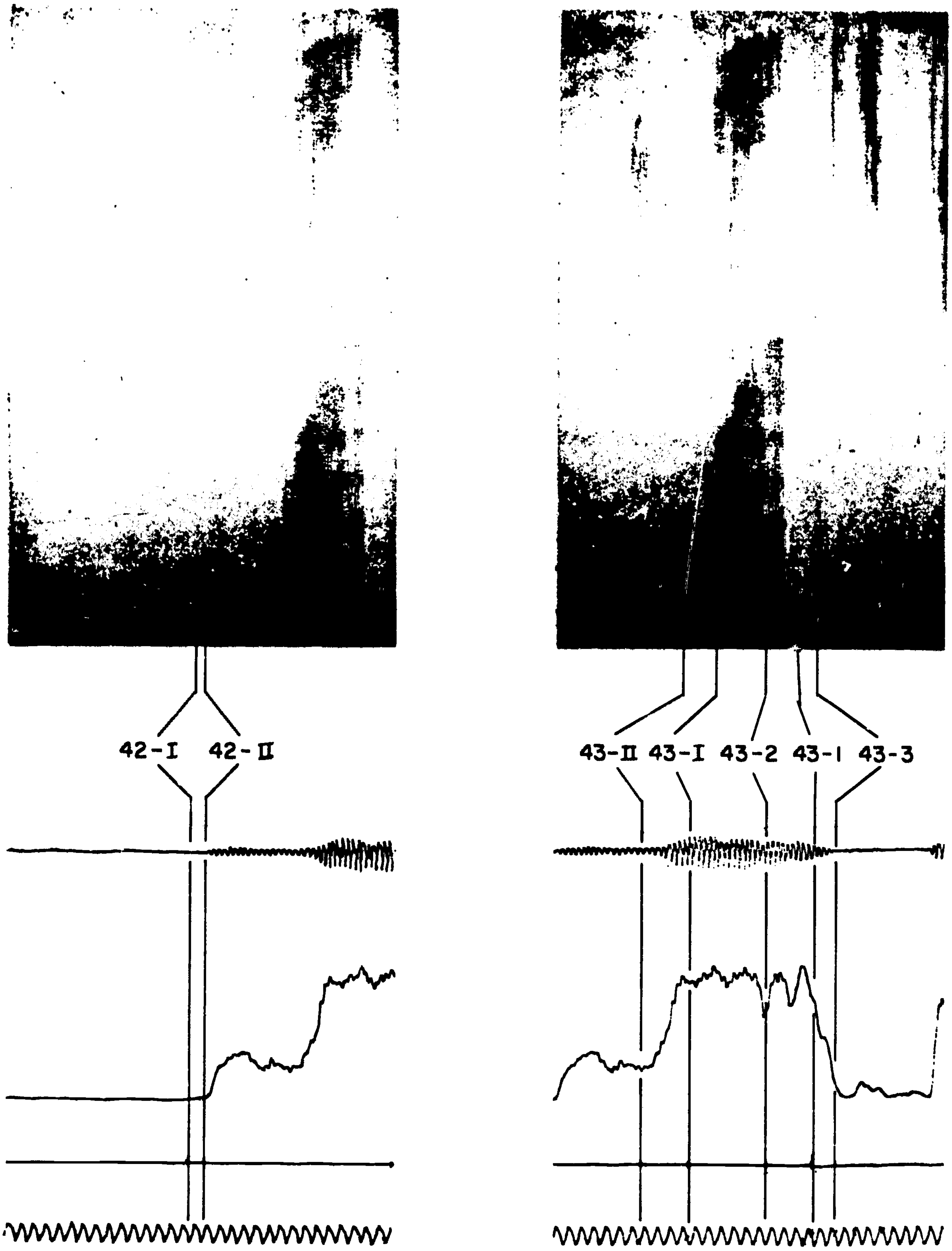

Fig. A36-A

Fig. A36-B

Fig. A36-A,B: Spectrogram and Mingogram Sections for Syllables 42 and 43

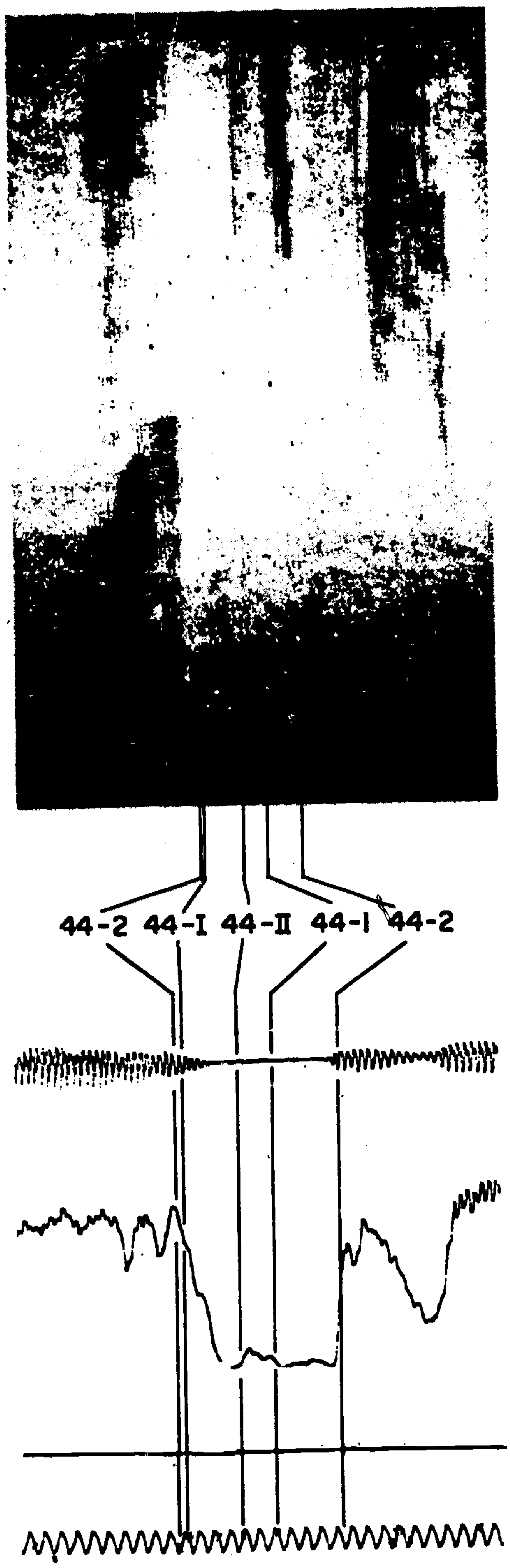

Fig. A37-A

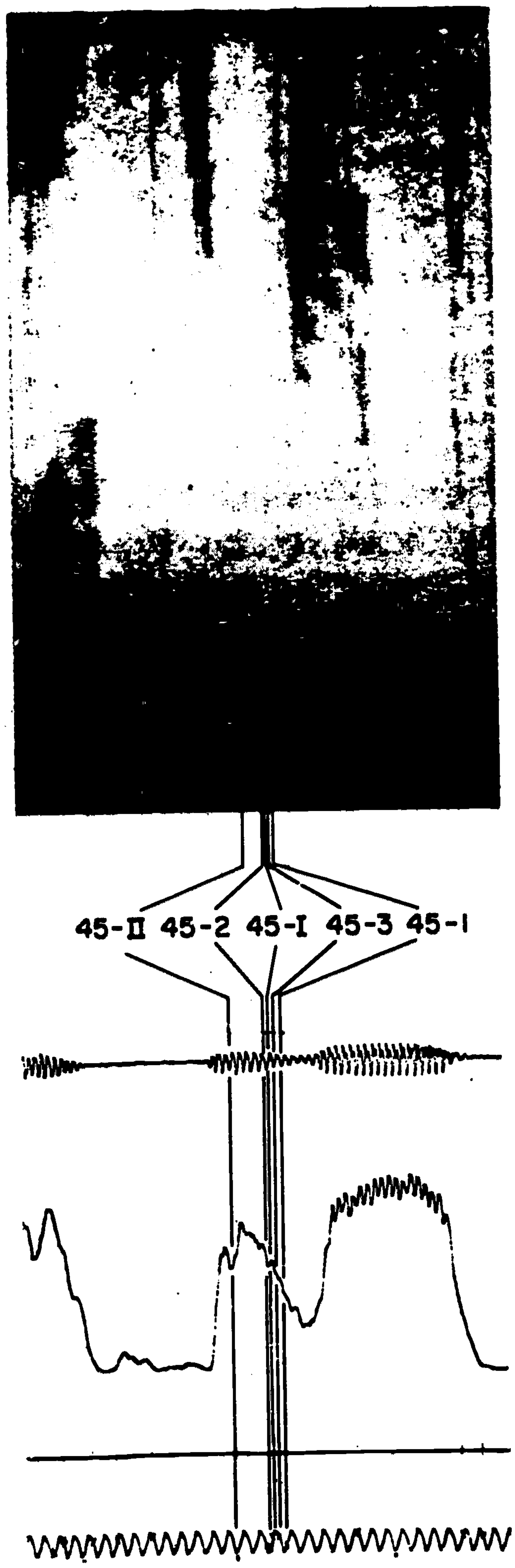

Fig. A37-B

Fig. A37-A,B: Spectrogram and Mingogram Sections
for Syllables 44 and 45

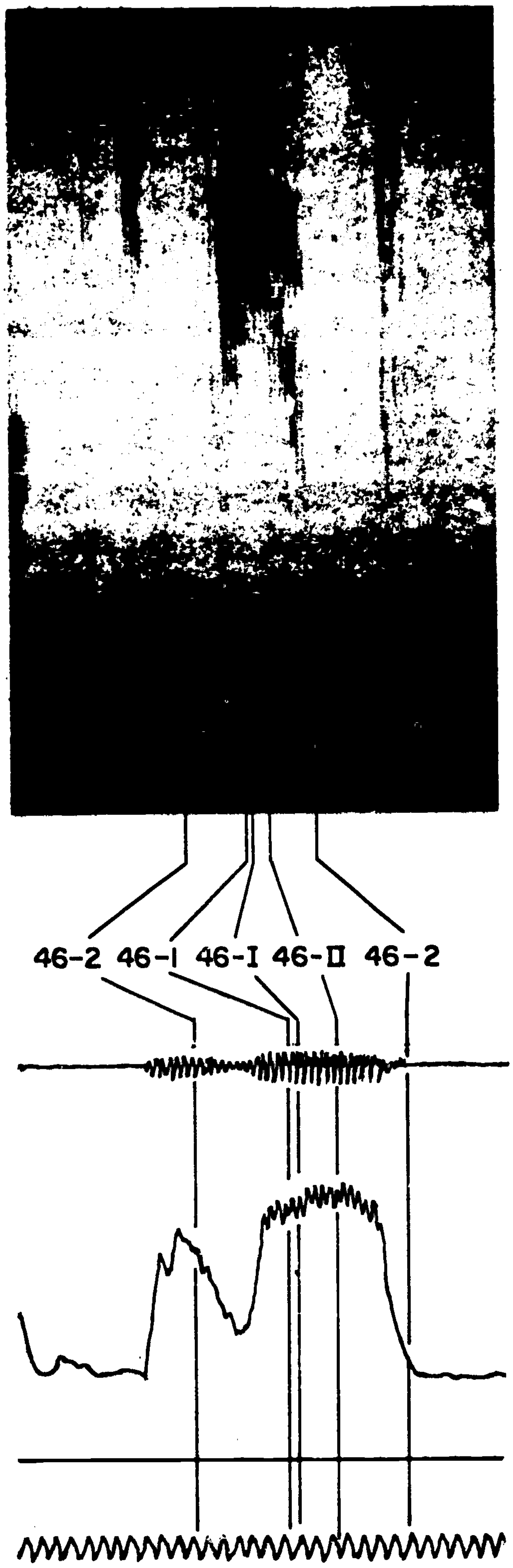

Fig. A38-A

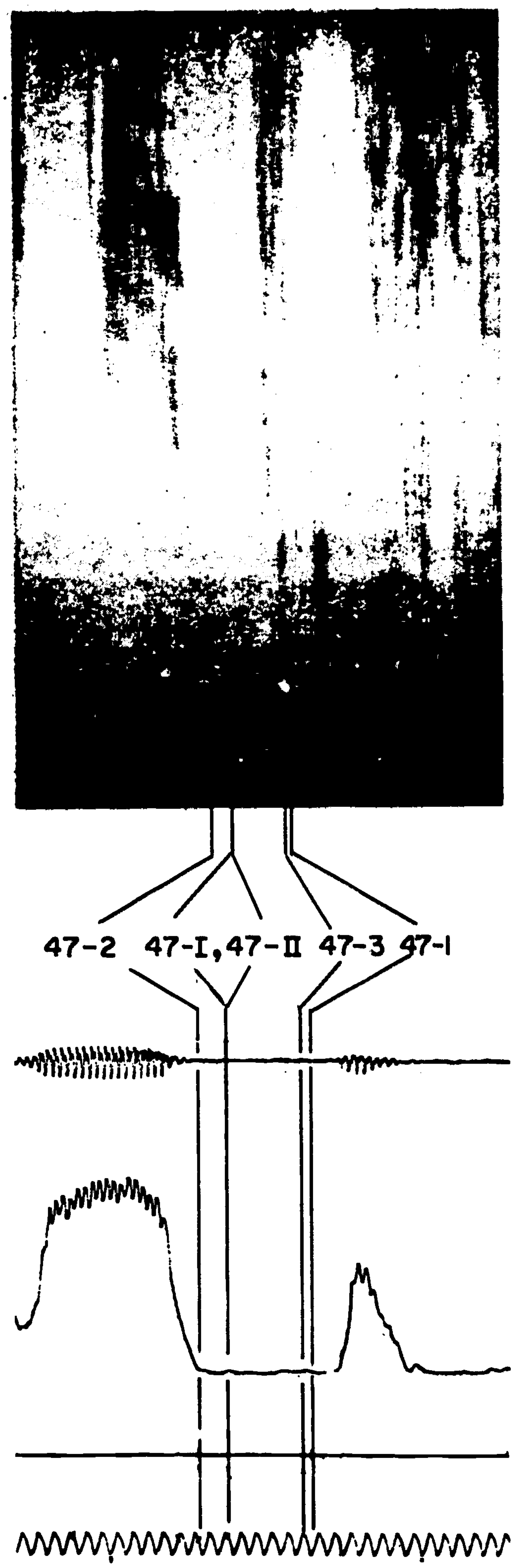

Fig. A38-B

Fig. A38-A,B: Spectrogram and Mingogram Sections for Syllables 46 and 47

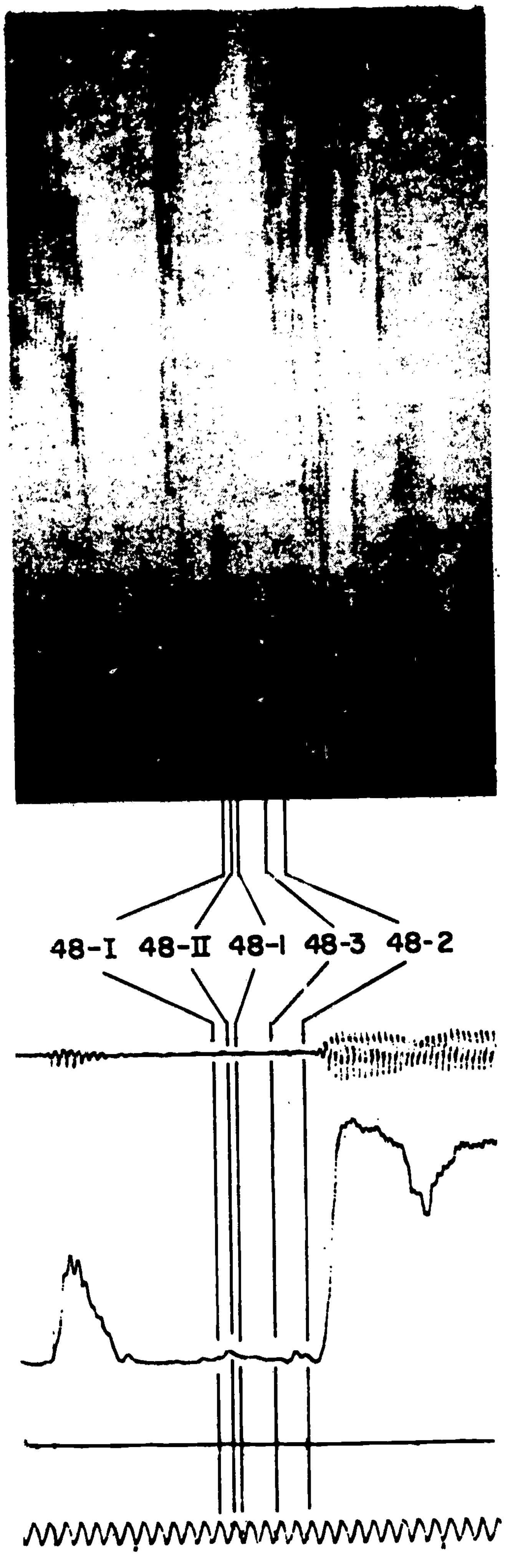

Fig. A39-A

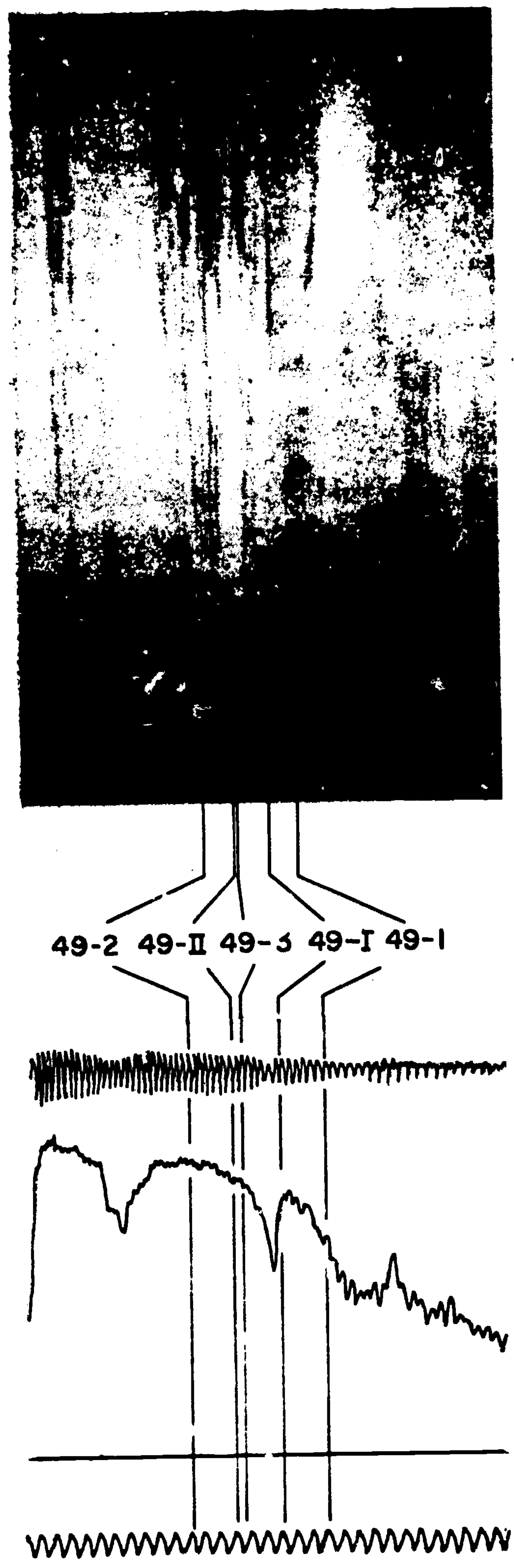

Fig. A39-B

Fig. A39-A,B: Spectrogram and Mingogram Sections for Syllables 48 and 49

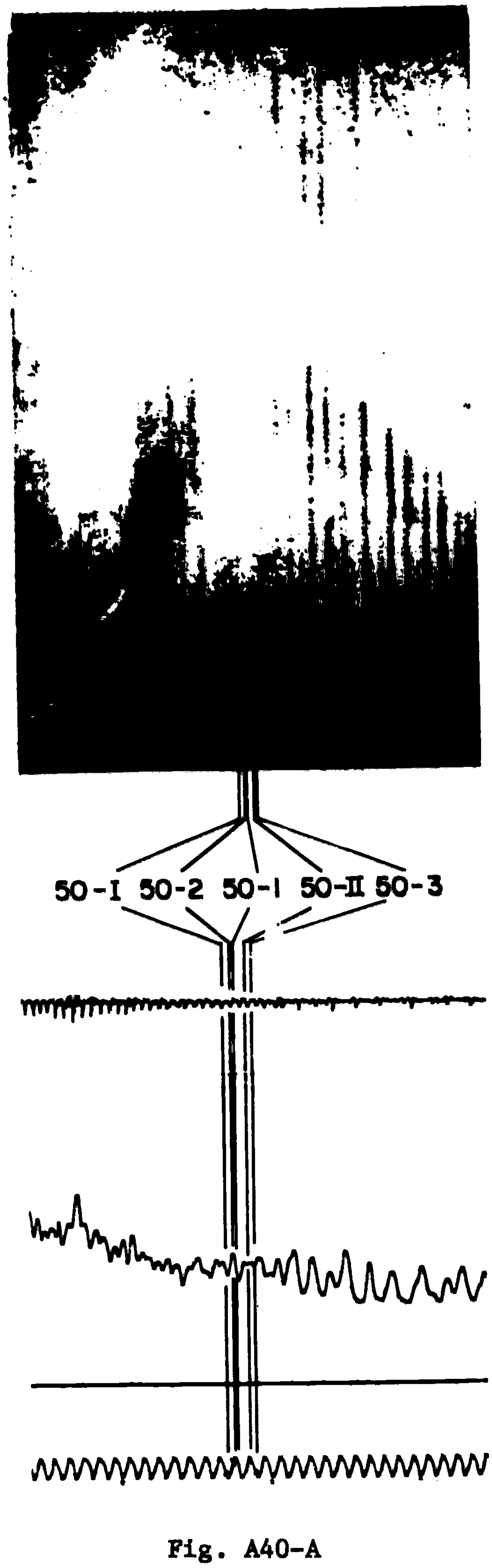

Fig. A40-A

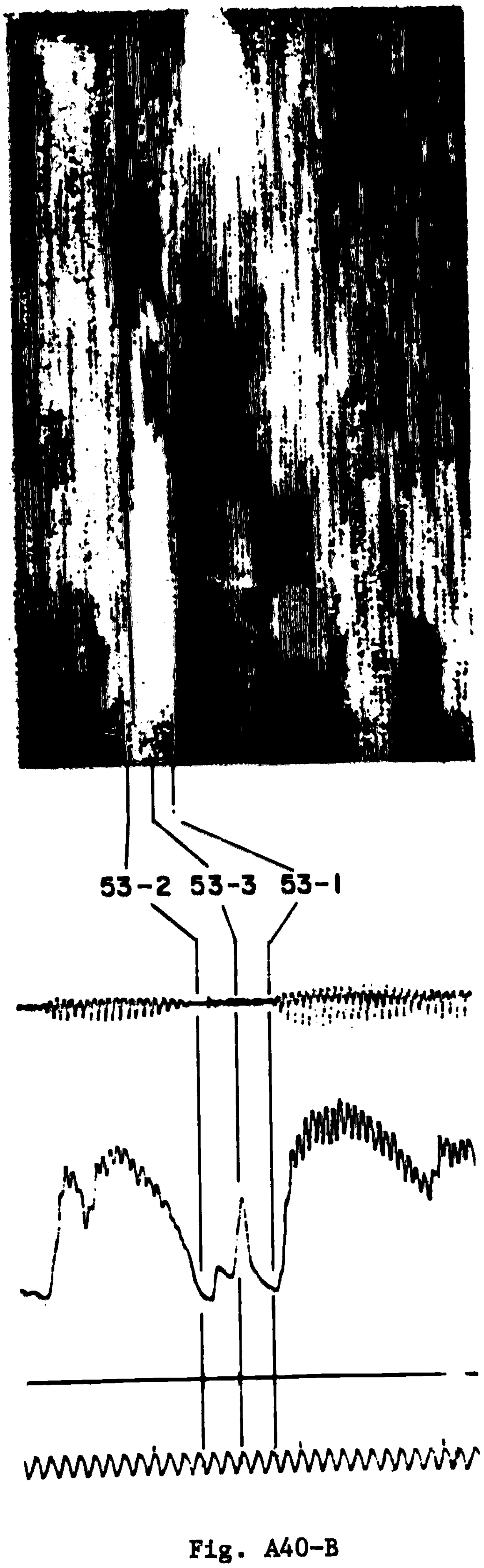

Fig. A40-B

Fig. A40-A,B: Spectrogram and Mingogram Sections
for Syllables 50 and 53

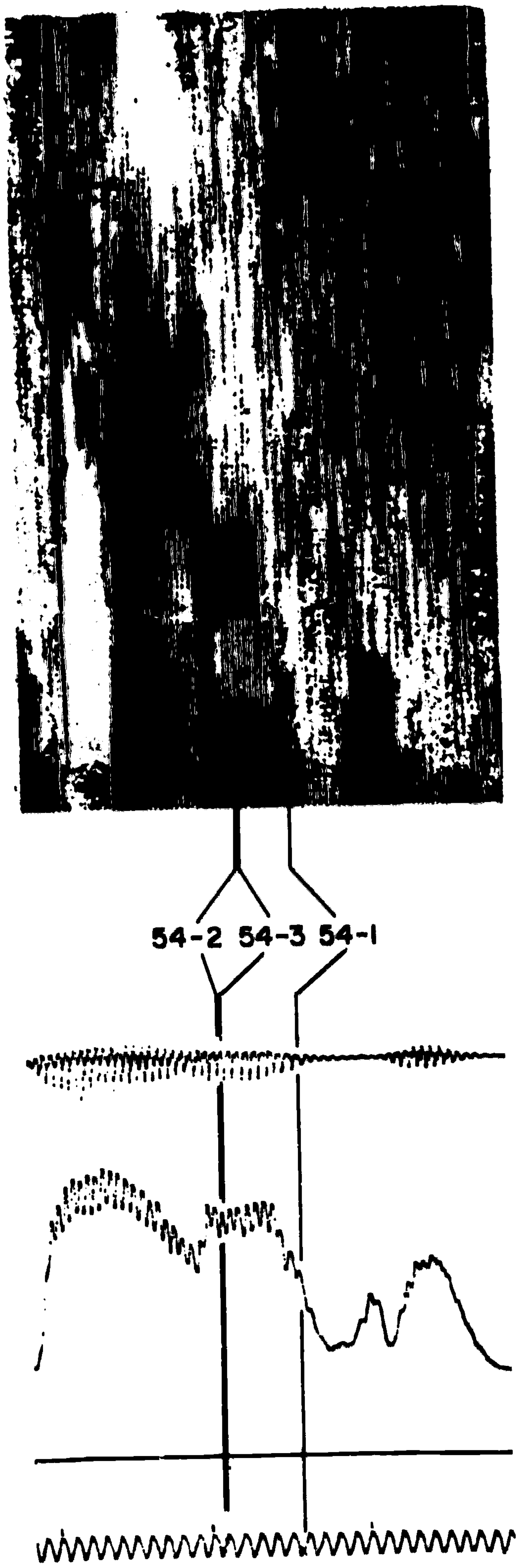

Fig. A41-A

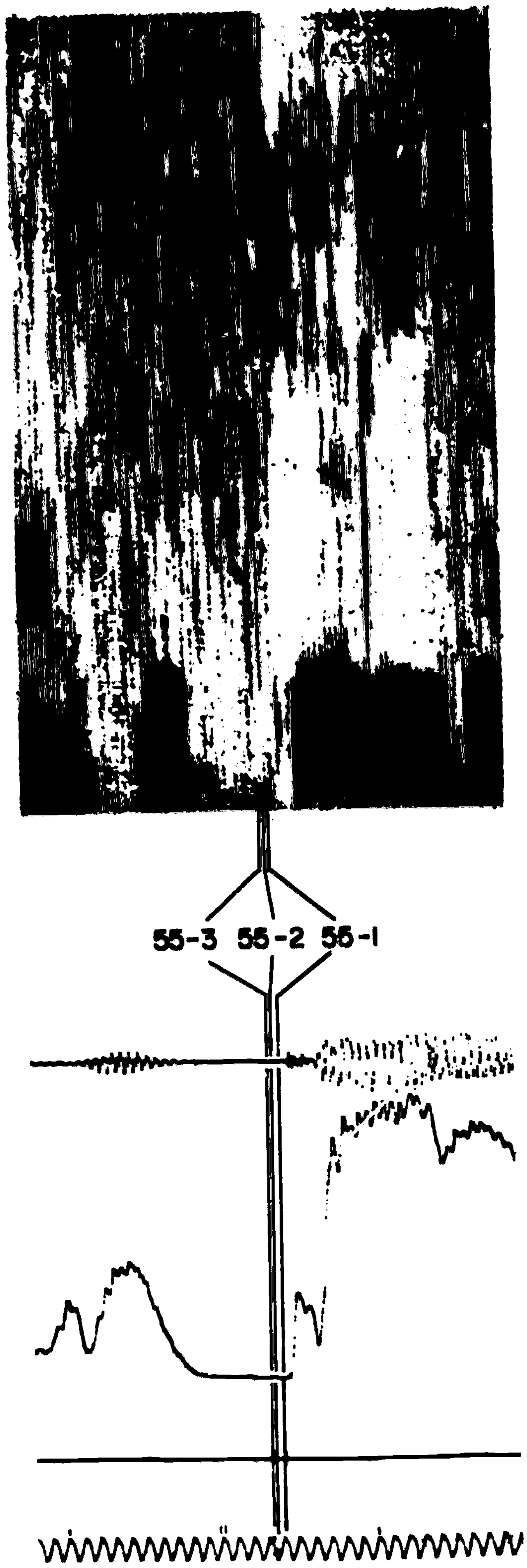

Fig. A41-B

Fig. A41-A,B: Spectrogram and Mingogram Sections for Syllables 54 and 55

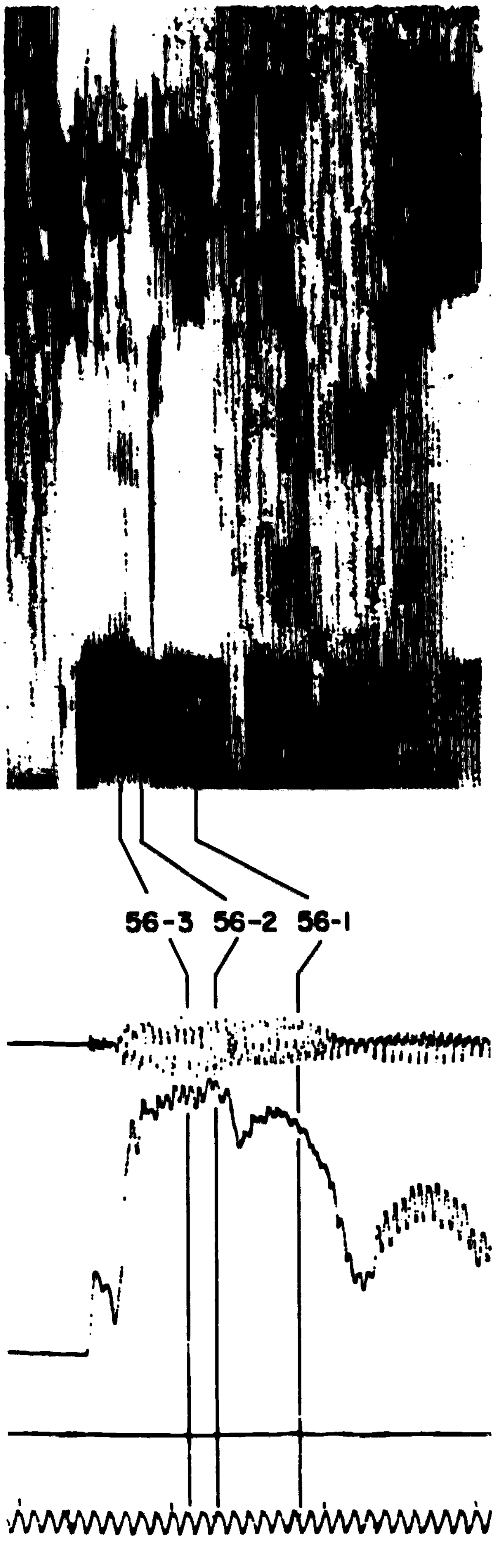

Fig. A42-A

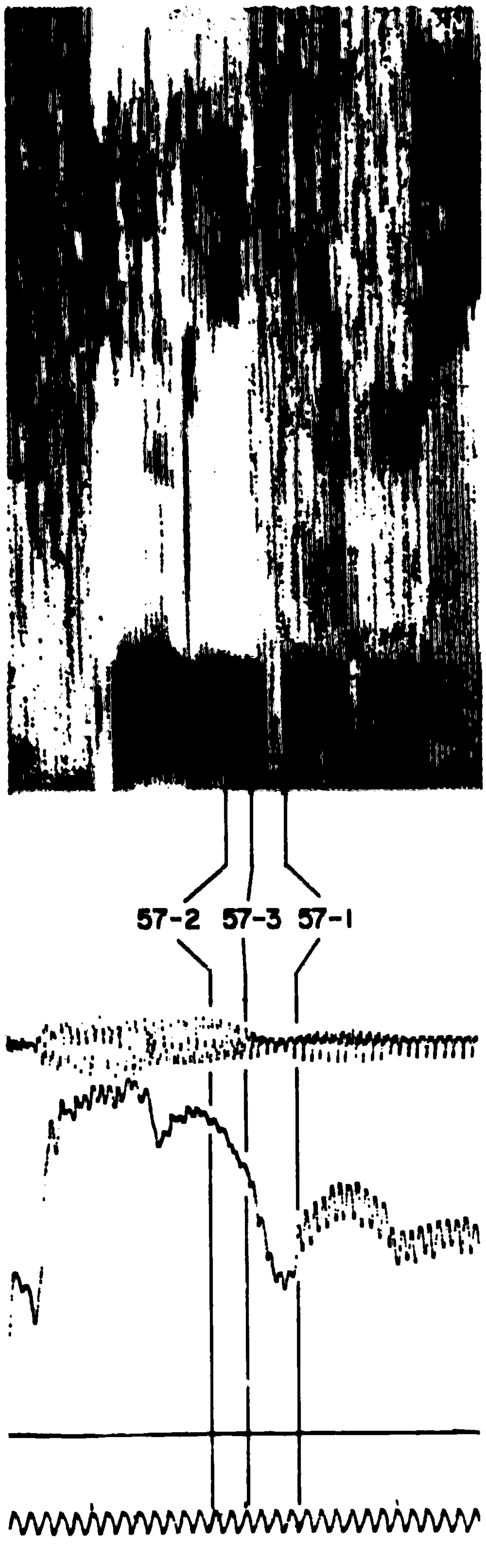

Fig. A42-B

Fig. A42-A,B: Spectrogram and Mingogram Sections for Syllables 56 and 57

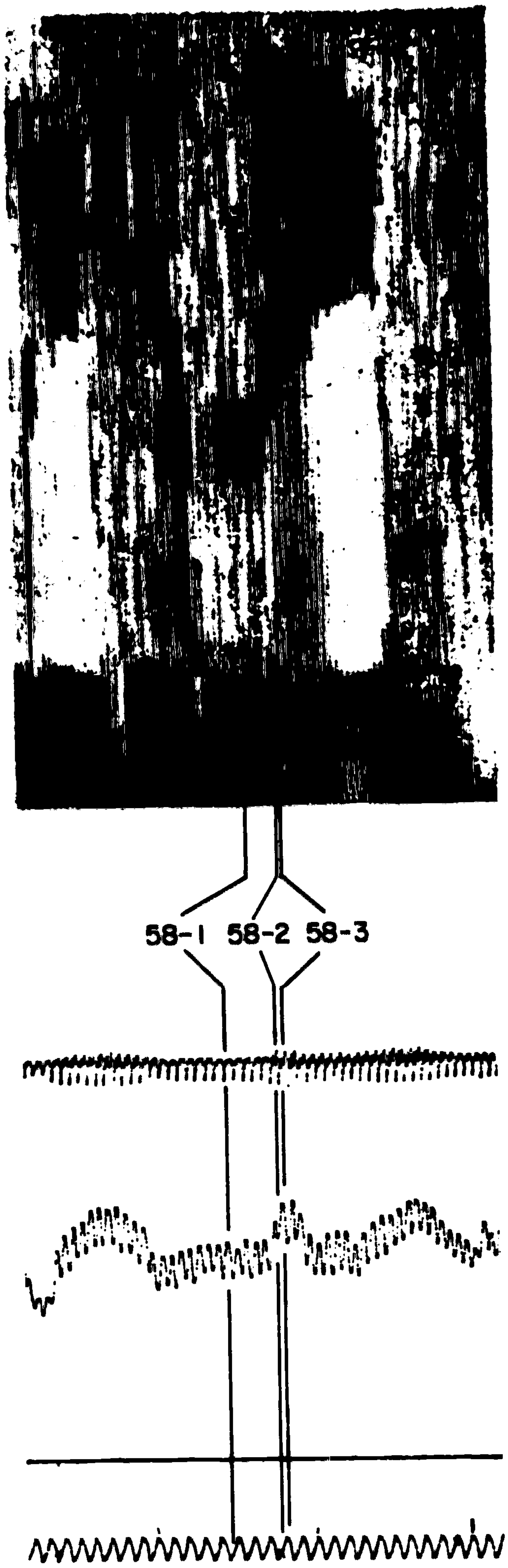

Fig. A43-A

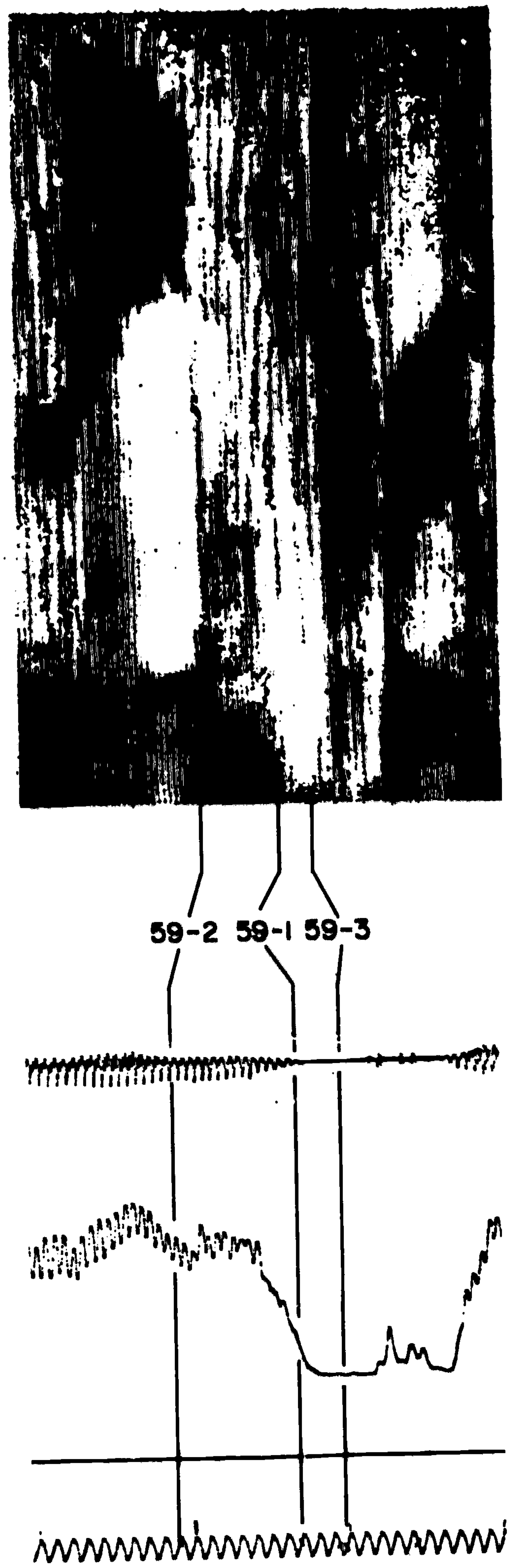

Fig. A43-B

Fig. A43-A,B: Spectrogram and Mingogram Sections for Syllables 58 and 59

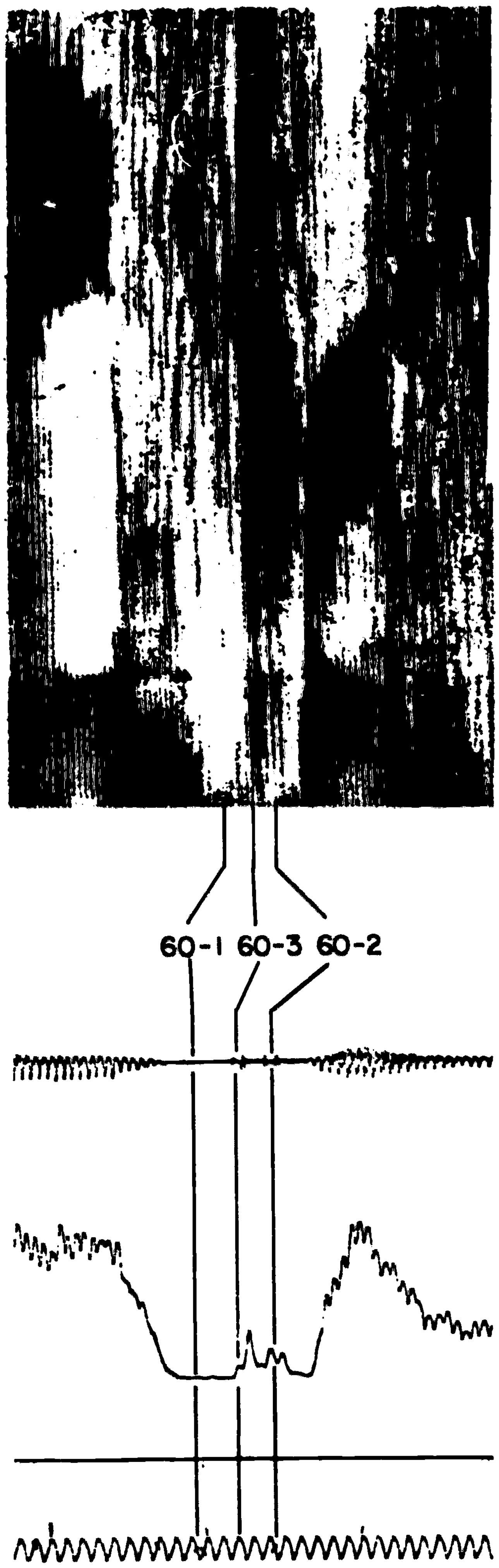

Fig. A44-A

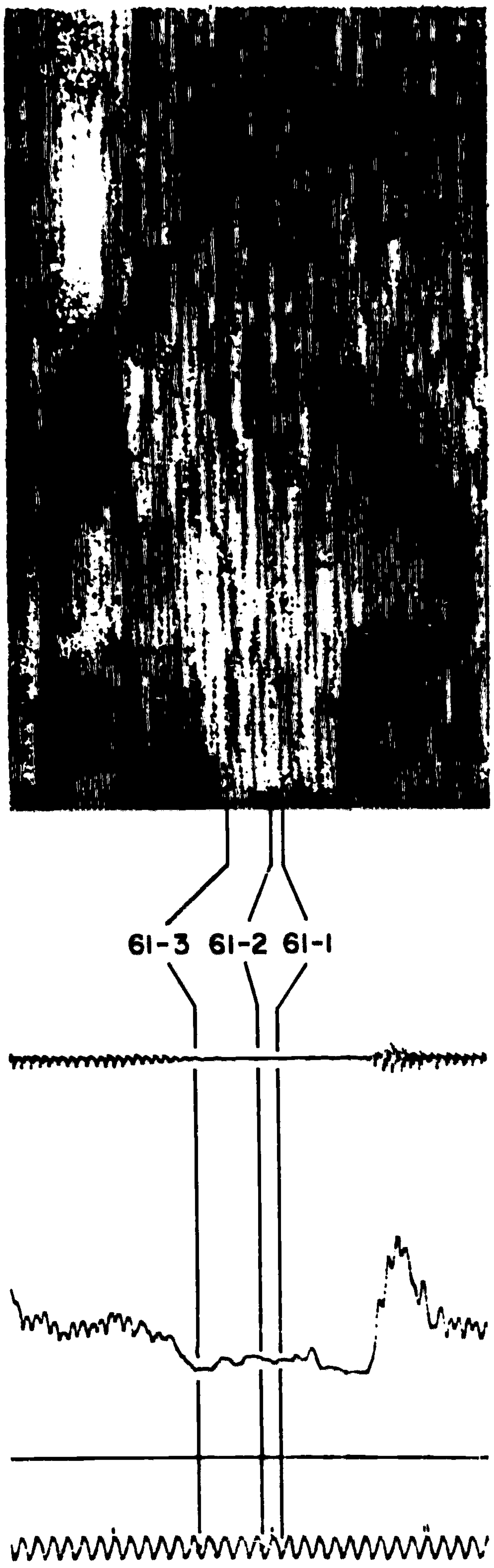

Fig. A44-B

Fig. A44-A,B: Spectrogram and Mingogram Sections for Syllables 60 and 61

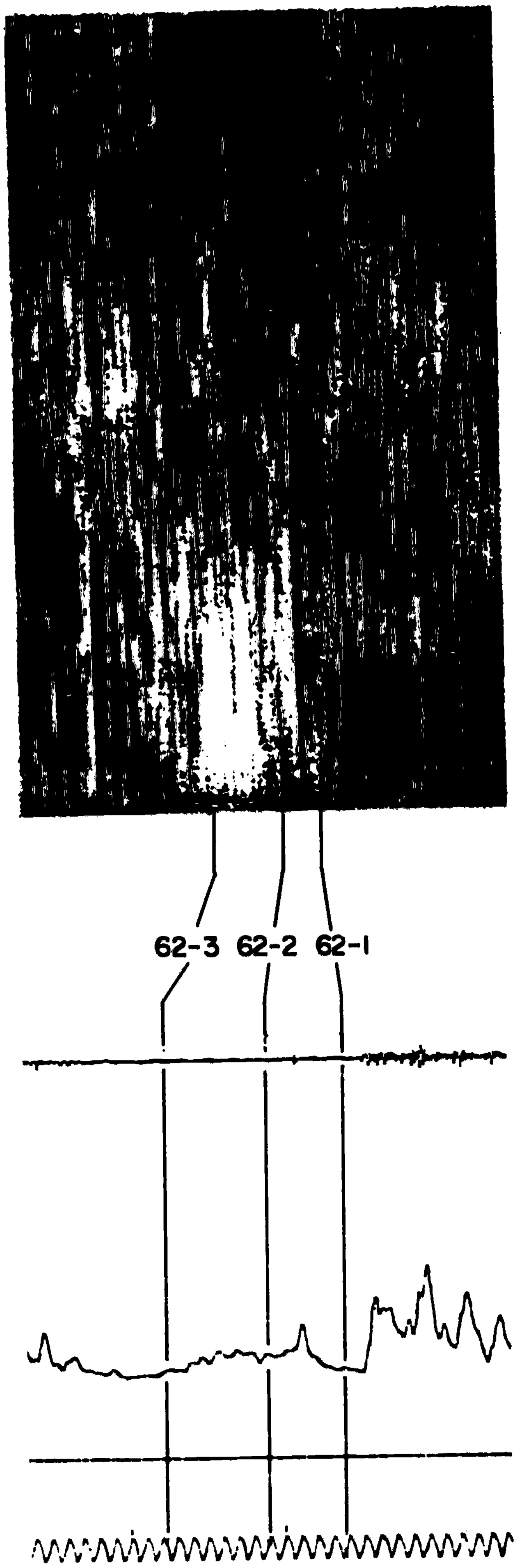

Fig. A45-A

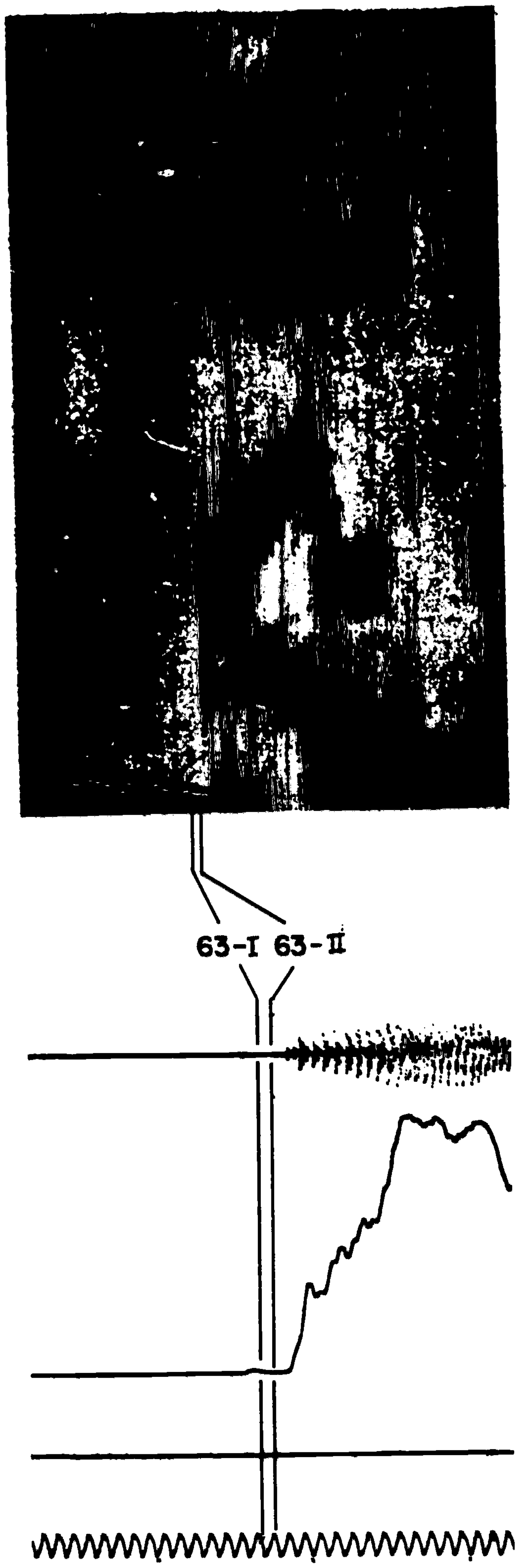

Fig. A45-B

Fig. A45-A,B: Spectrogram and Mingogram Sections for Syllables 62 and 63

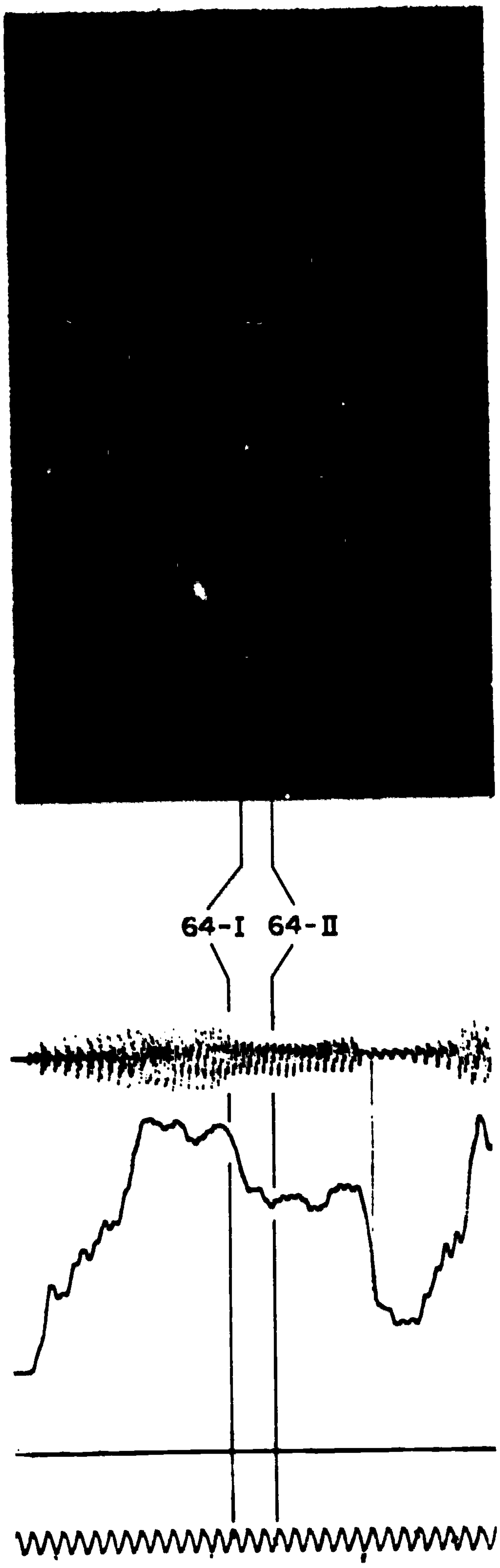

Fig. A46-A

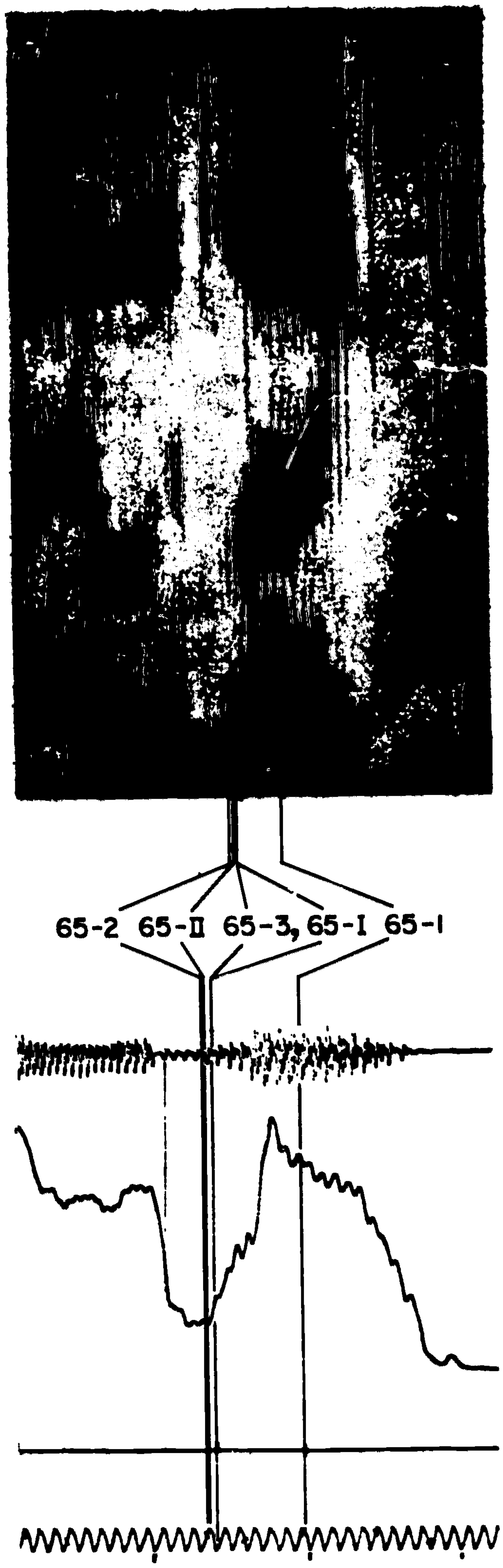

Fig. A46-B

Fig. A46-A,B: Spectrogram and Mingogram Sections for Syllables 64 and 65

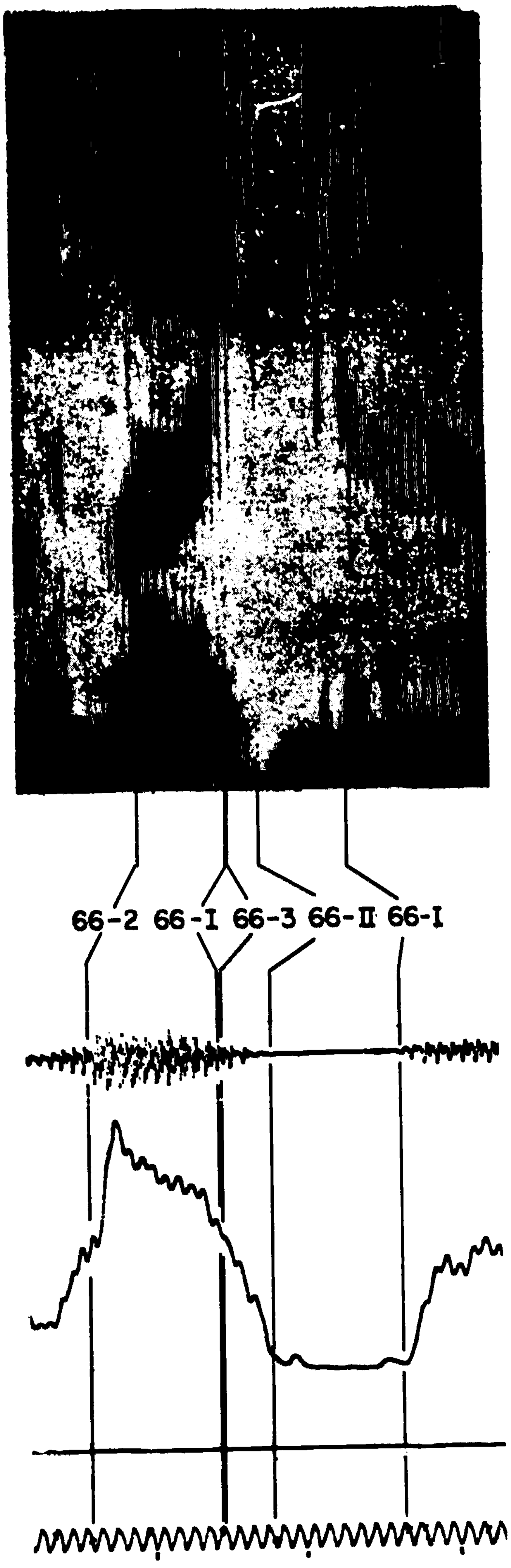

Fig. A47-A

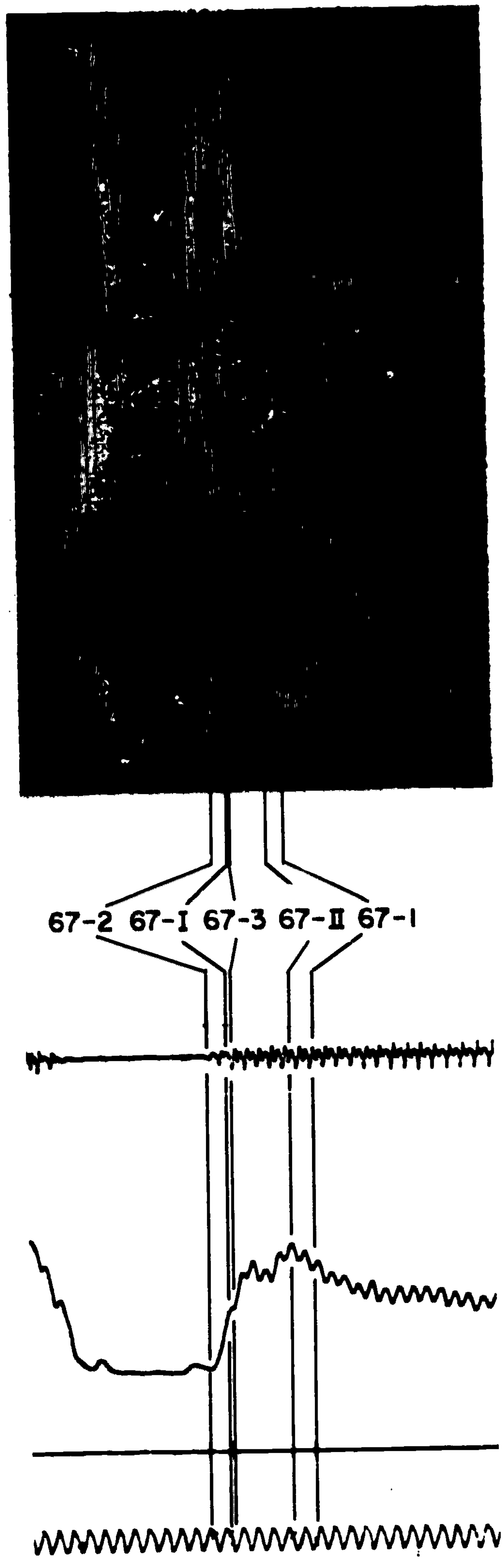

Fig. 47-B

Fig. A47-A,B: Spectrogram and Mingogram Sections
for Syllables 66 and 67

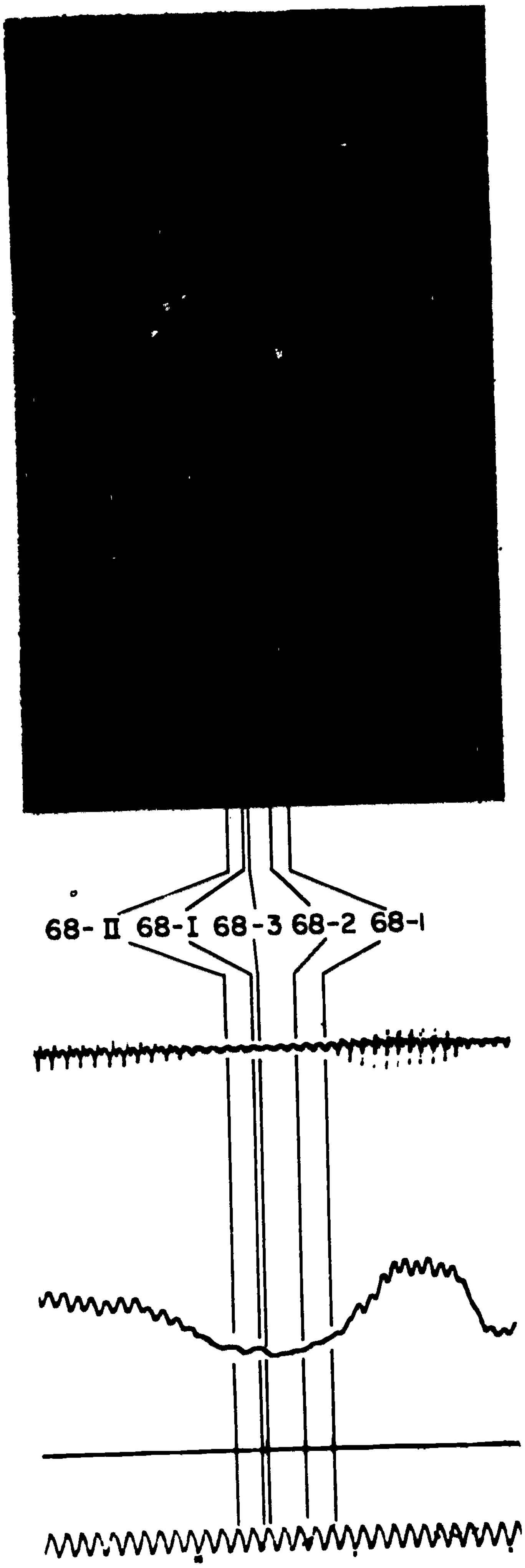

Fig. A48-A

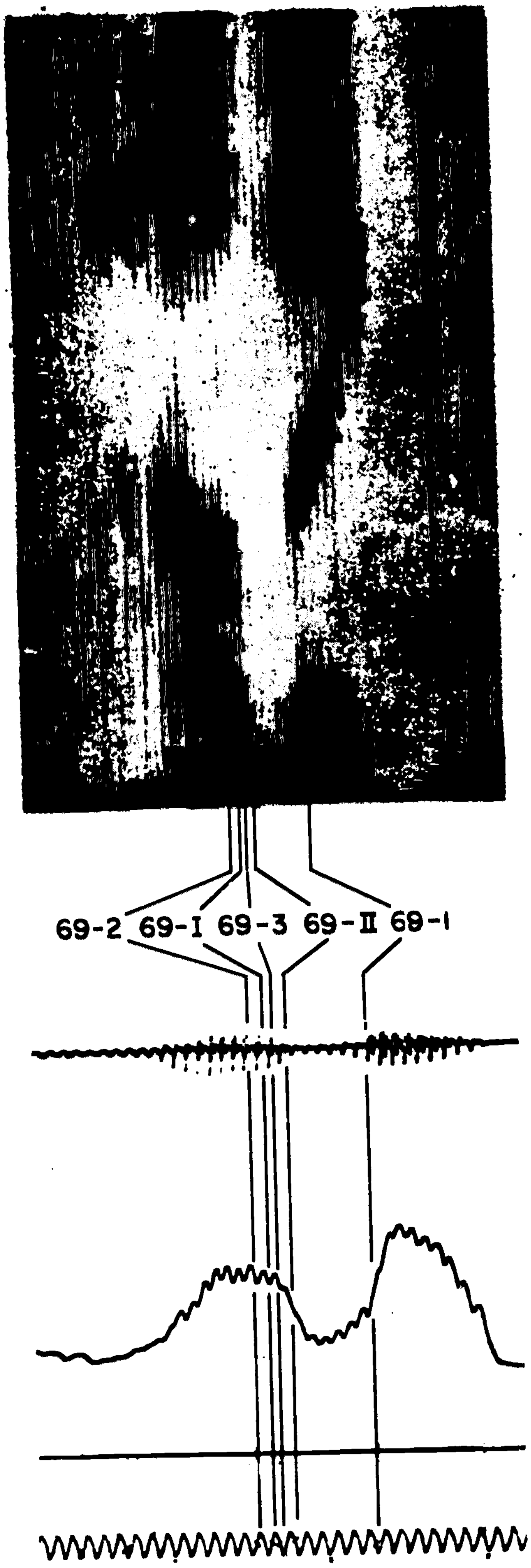

Fig. A48-B

Fig. A48-A,B: Spectrogram and Mingogram Sections
for Syllables 68 and 69

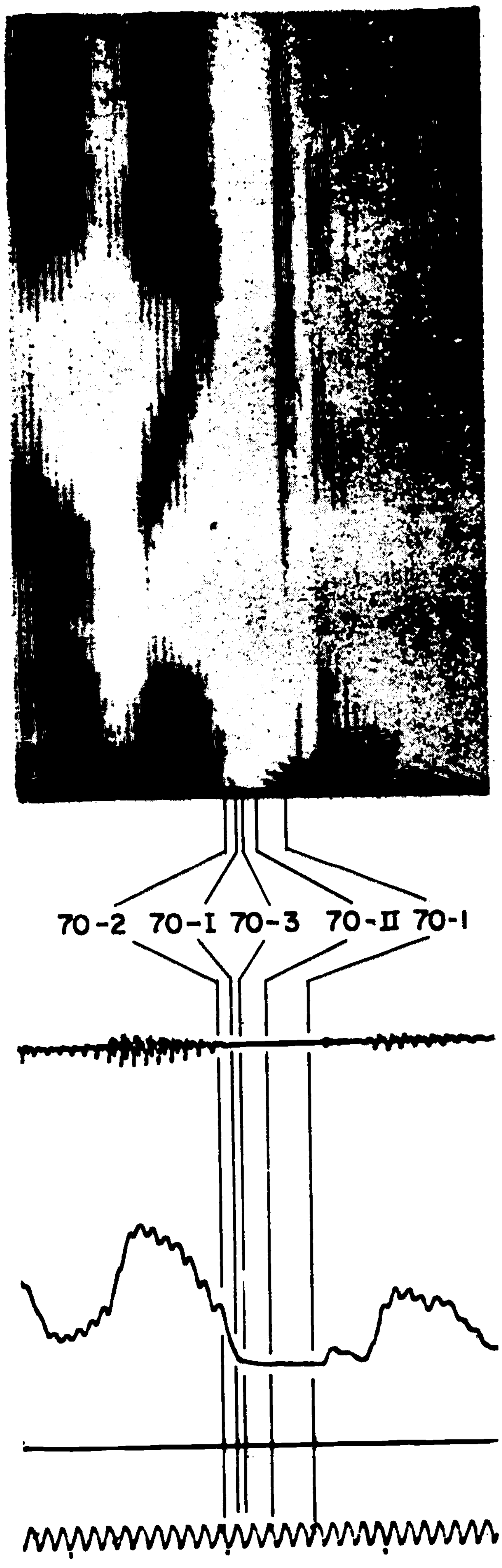

Fig. A49-A

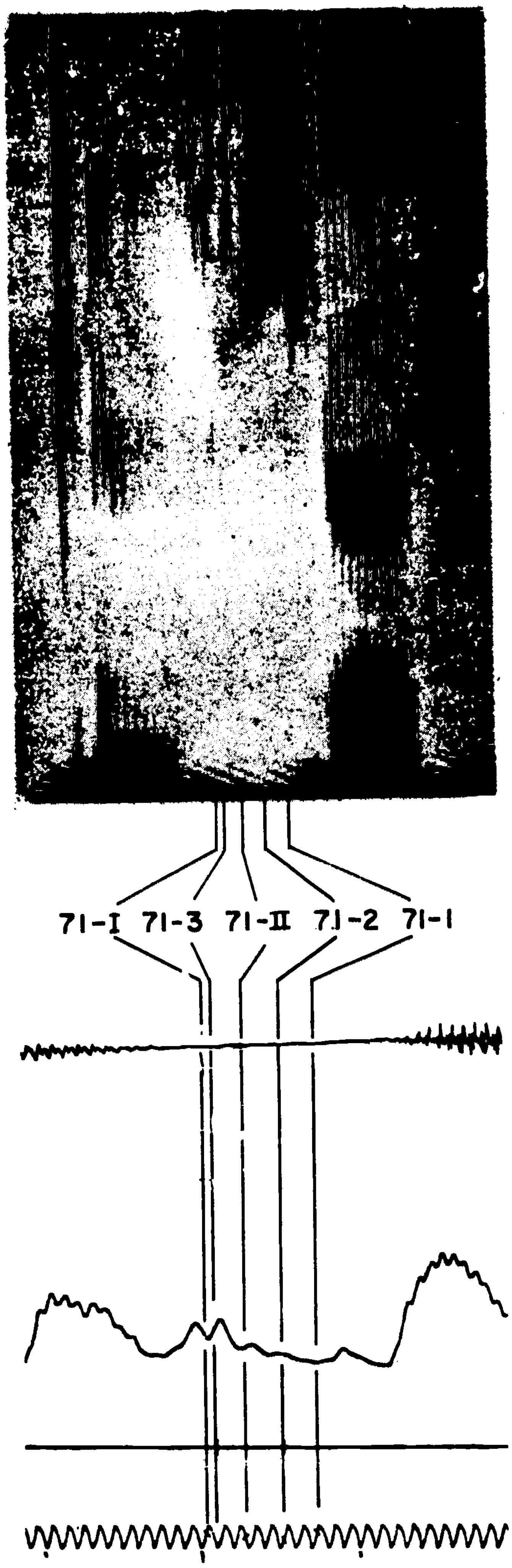

Fig. A49-B

Fig. A49-A,B: Spectrogram and Mingogram Sections for Syllables 70 and 71

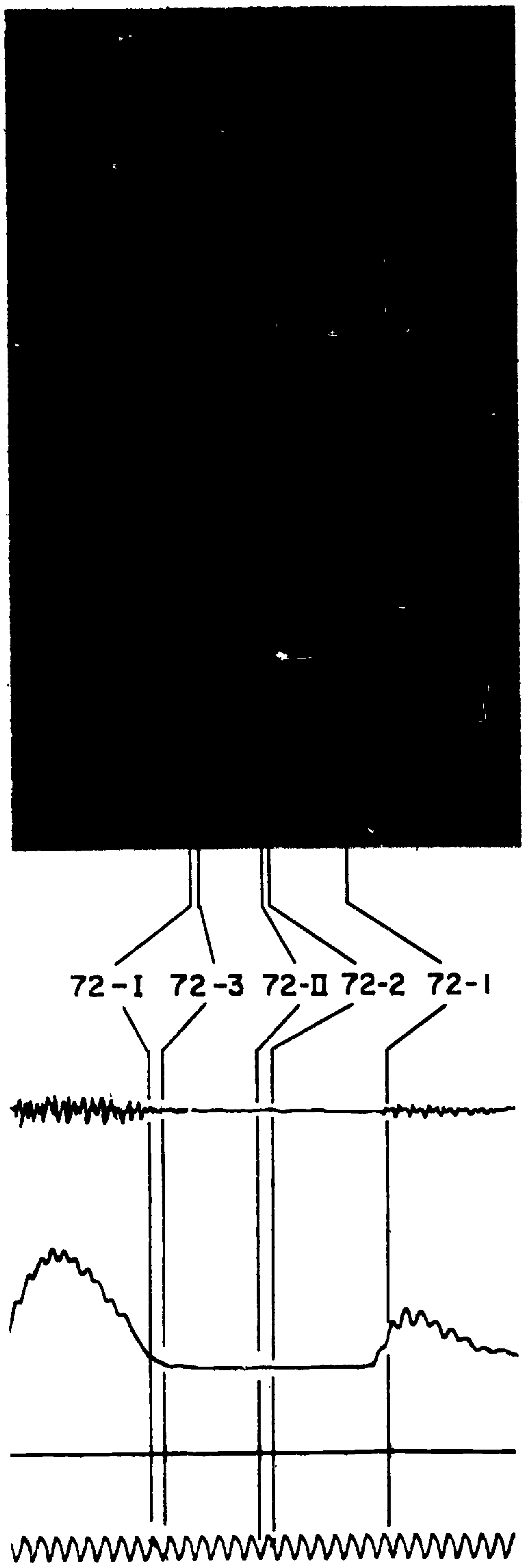

Fig. A50-A

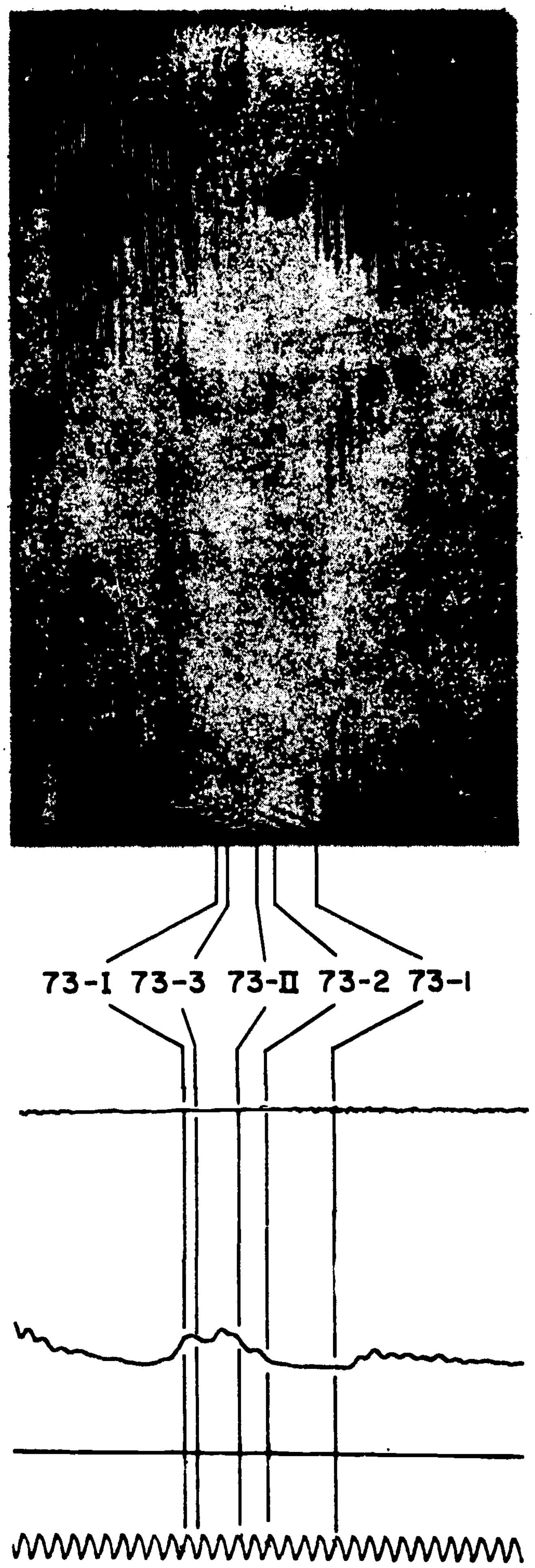

Fig. A50-B

Fig. A50-A,B: Spectrogram and Mingogram Sections
for Syllables 72 and 73

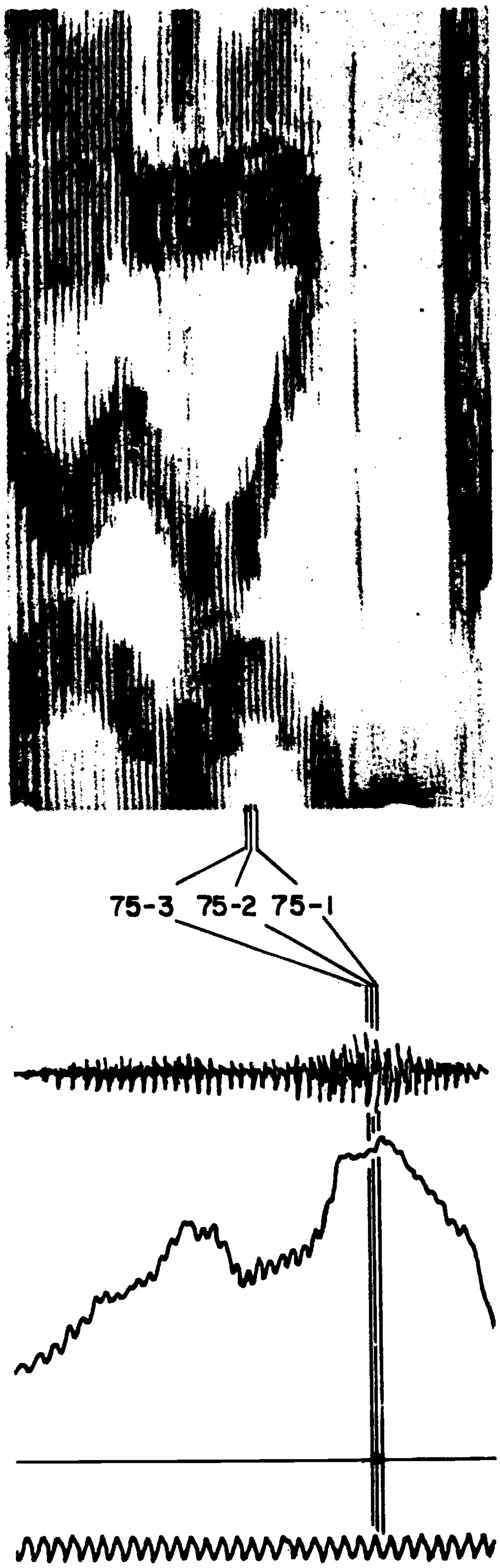

Fig. A51-A

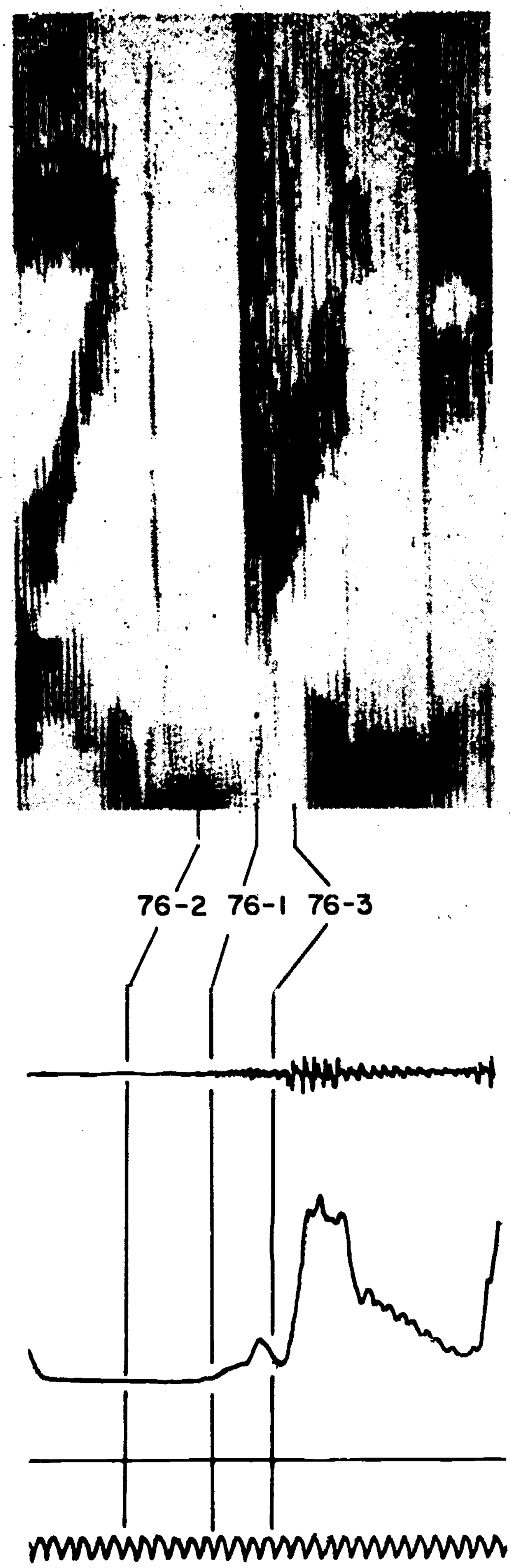

Fig. A51-B

Fig. A51-A,B: Spectrogram and Mingogram Sections for Syllables 75 and 76

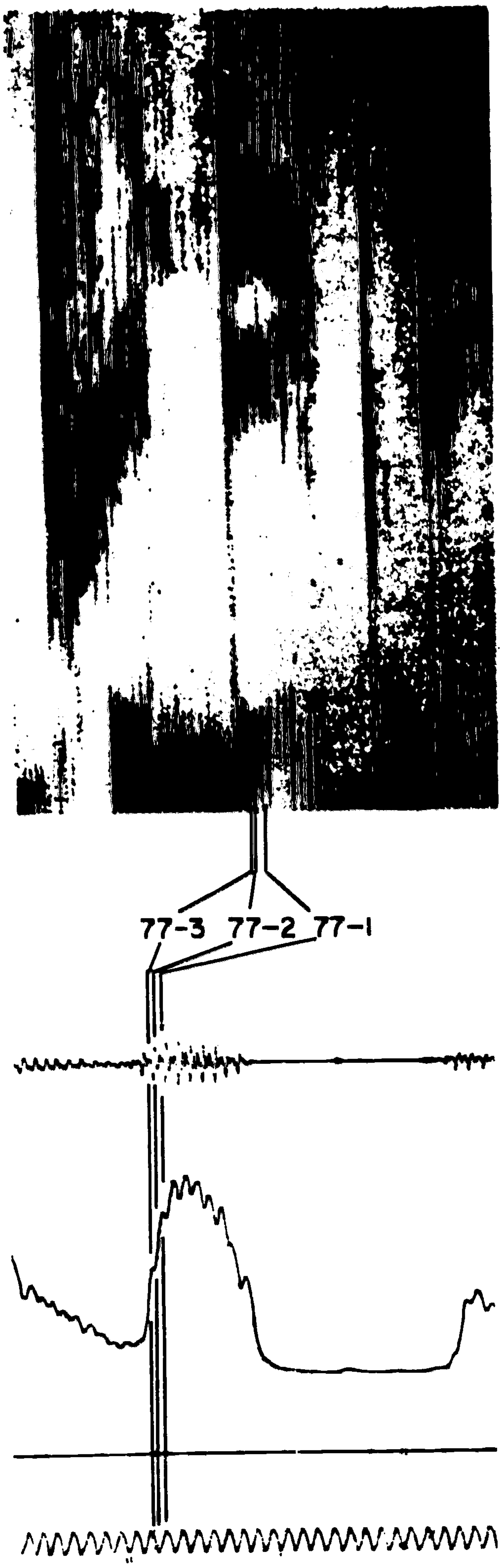

Fig. A52-A

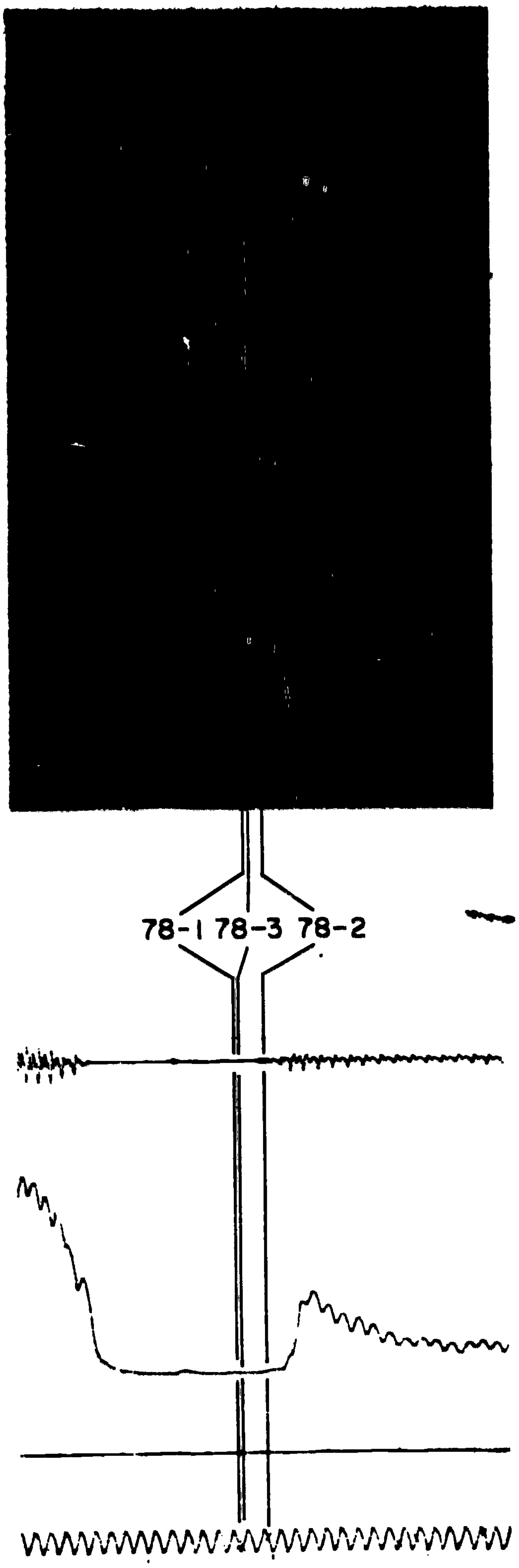

Fig. A52-B

Fig. A52-A,B: Spectrogram and Mingogram Sections
for Syllables 77 and 78

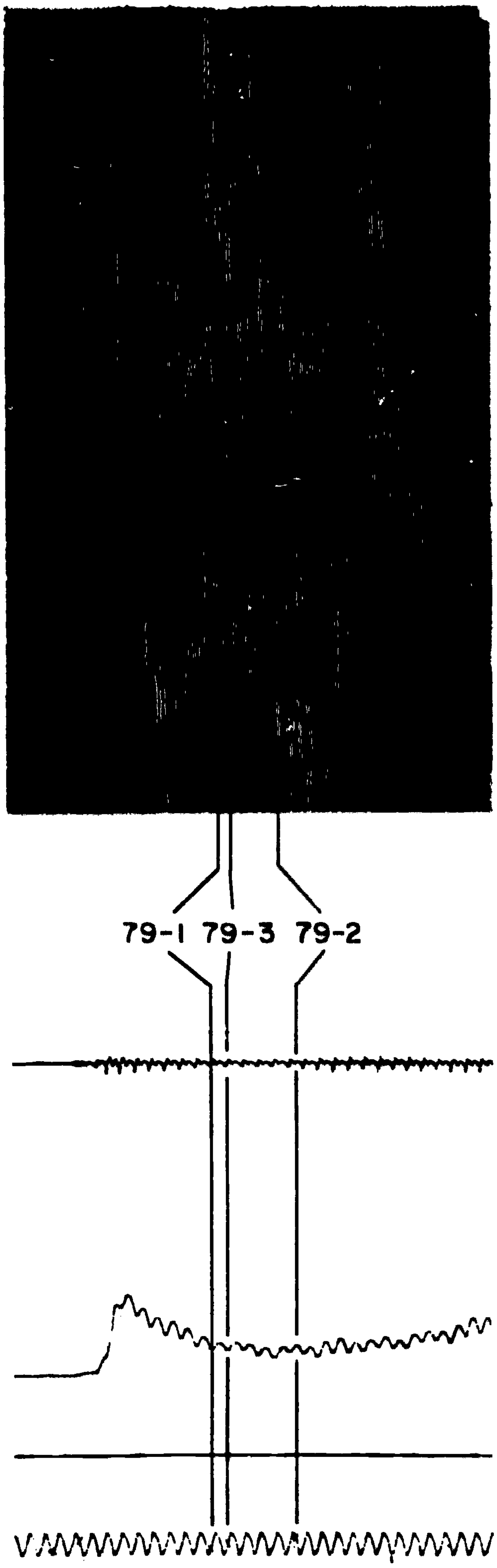

Fig. A53-A

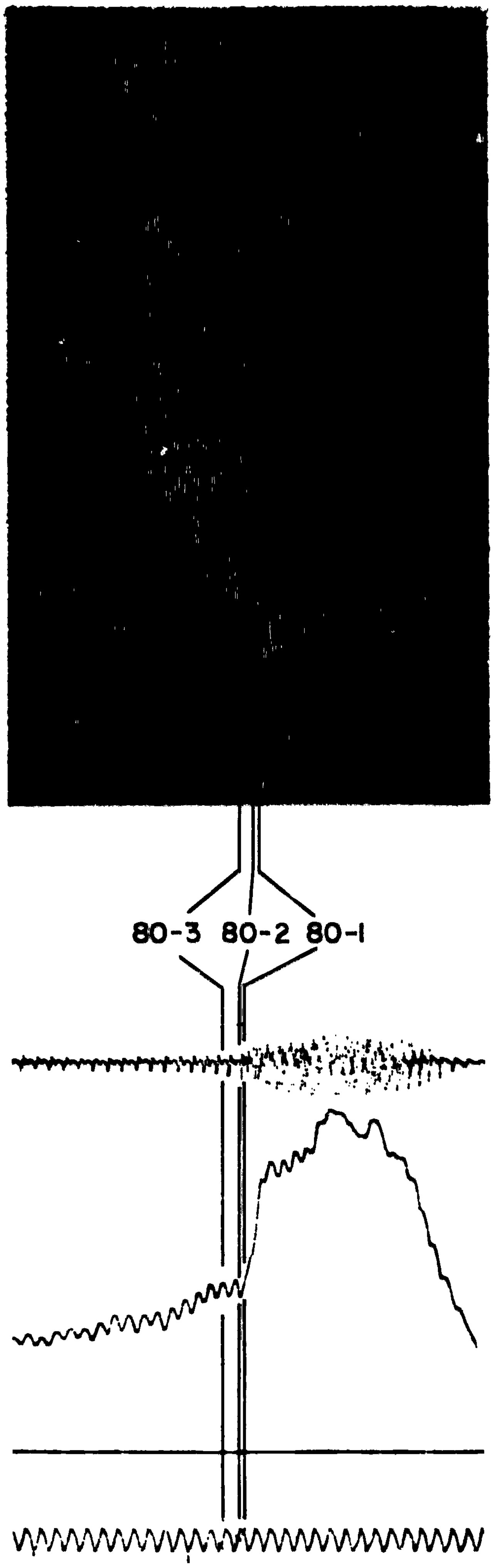

Fig. A53-B

Fig. A53-A,B: Spectrogram and Mingogram Sections for Syllables 79 and 80

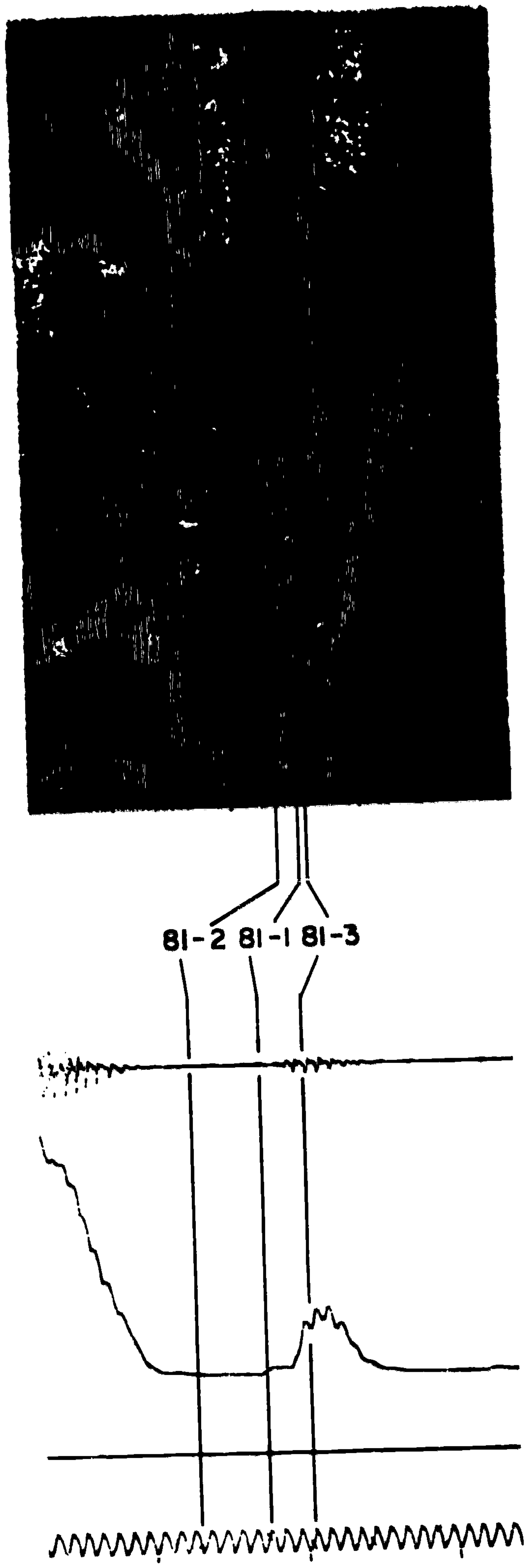

Fig. A54-A

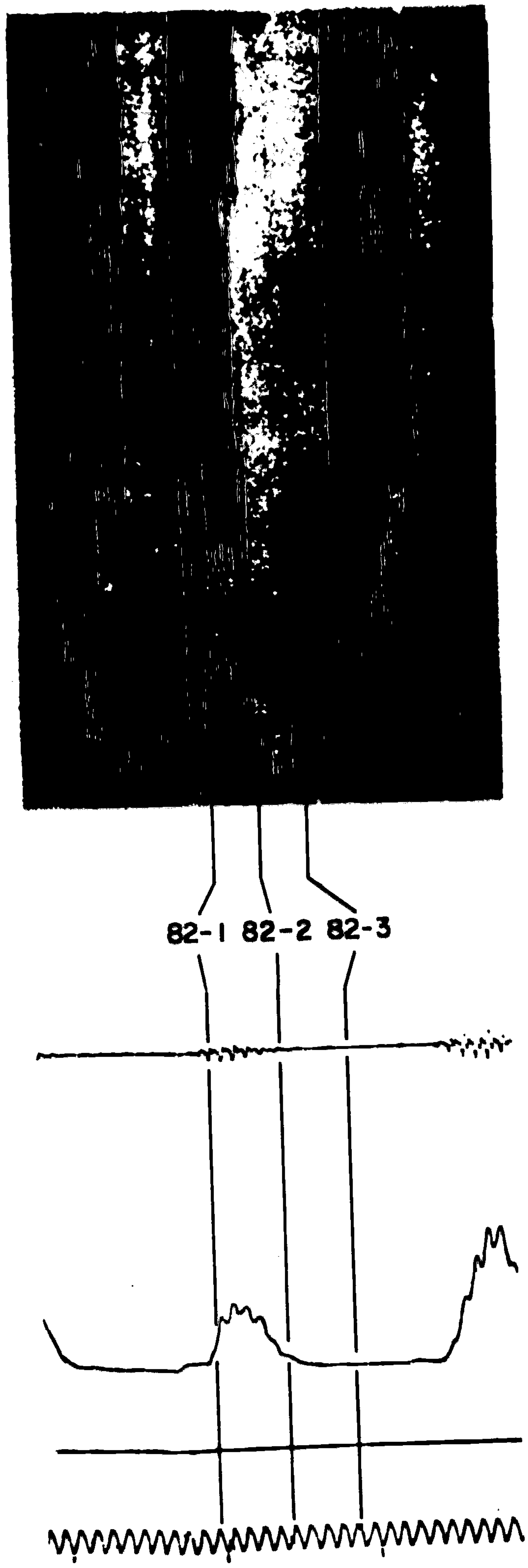

Fig. A54-B

Fig. A54-A,B: Spectrogram and Mingogram Sections for Syllables 81 and 82

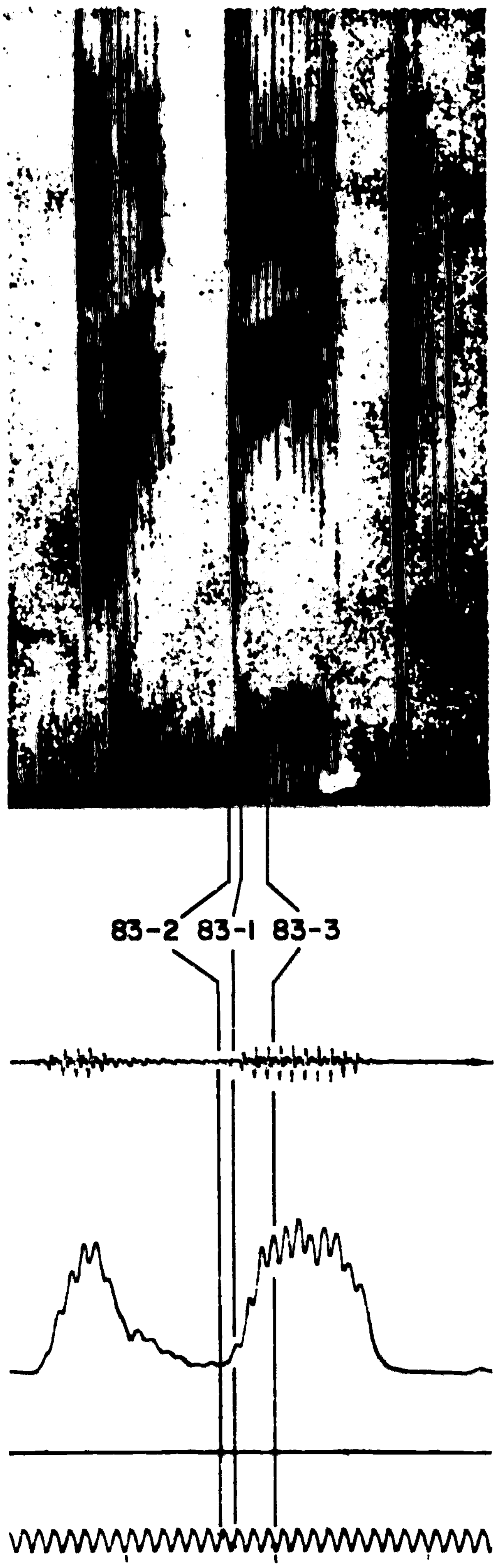

Fig. A55-A

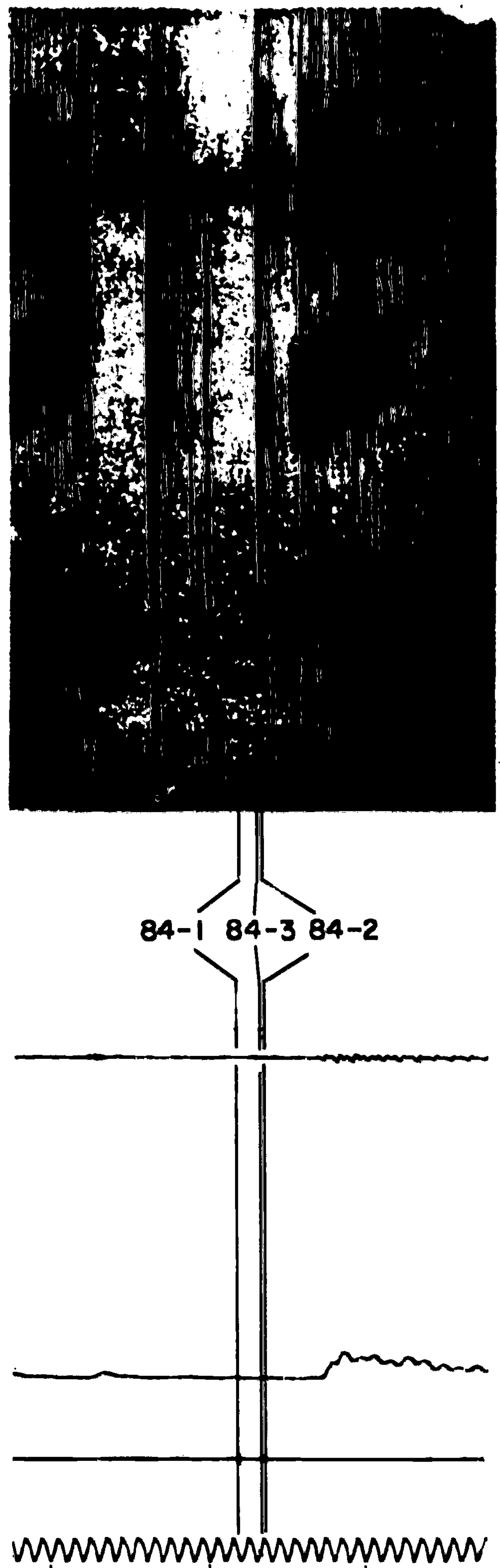

Fig. A55-B

Fig. A55-A,B: Spectrogram and Mingogram Sections for Syllables 83 and 84

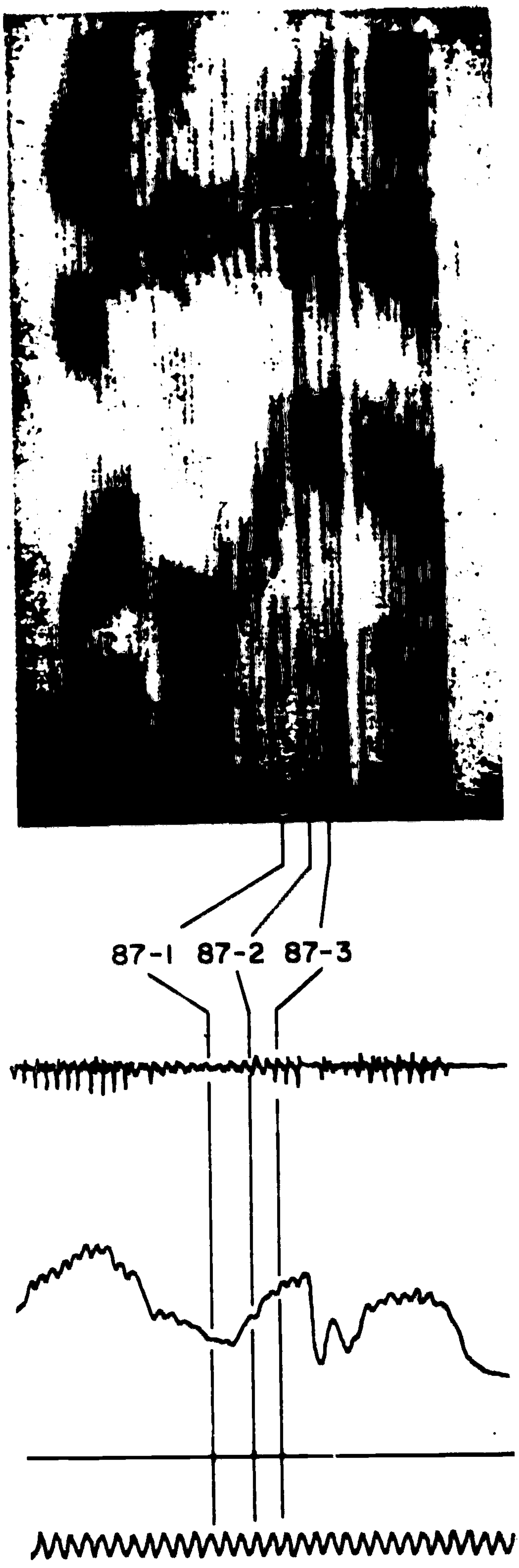

Fig. A56-A

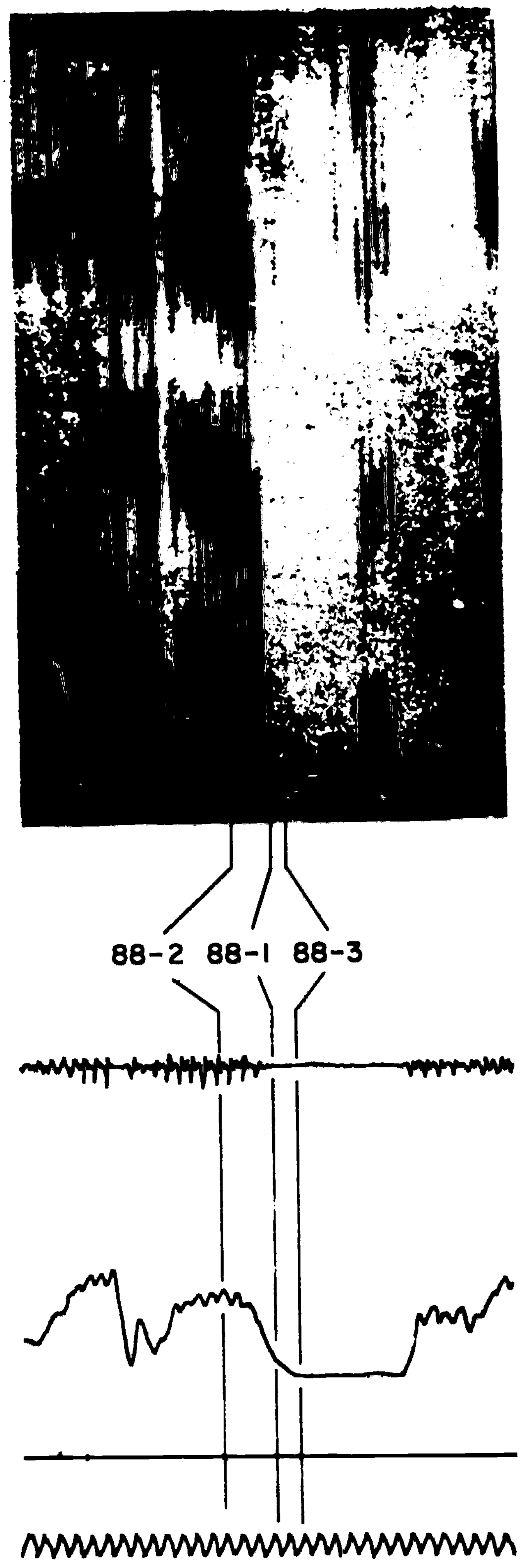

Fig. A56-B

Fig. A56-A,B: Spectrogram and Mingogram Sections for Syllables 87 and 88

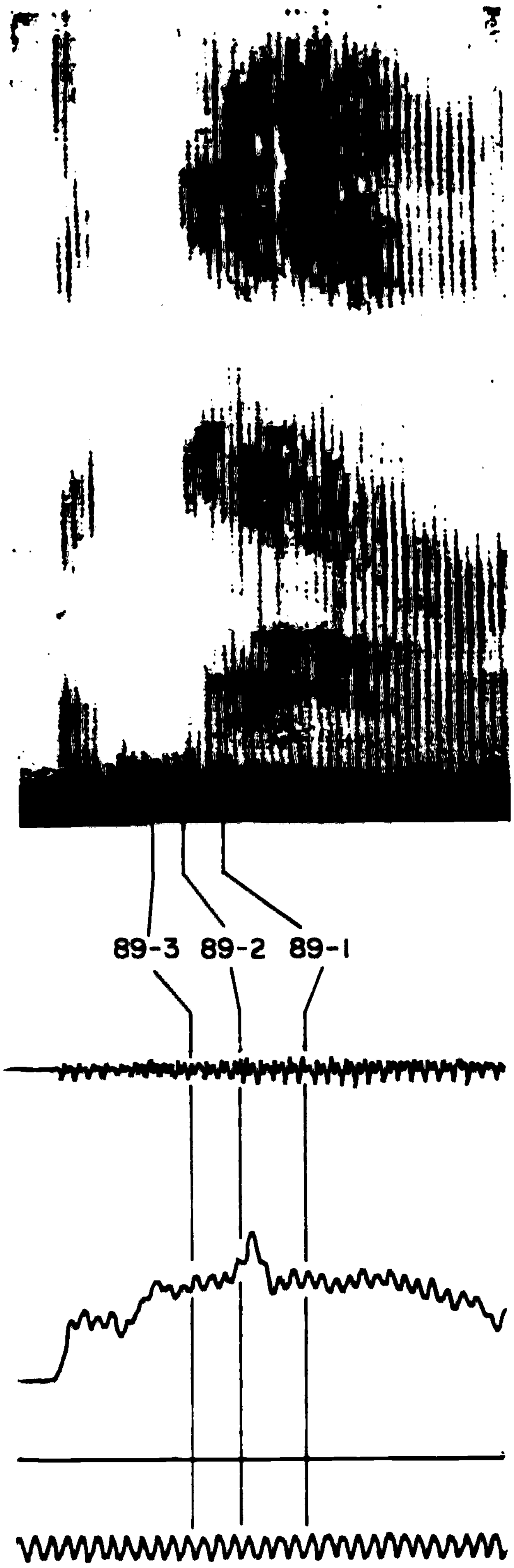

Fig. A57-A

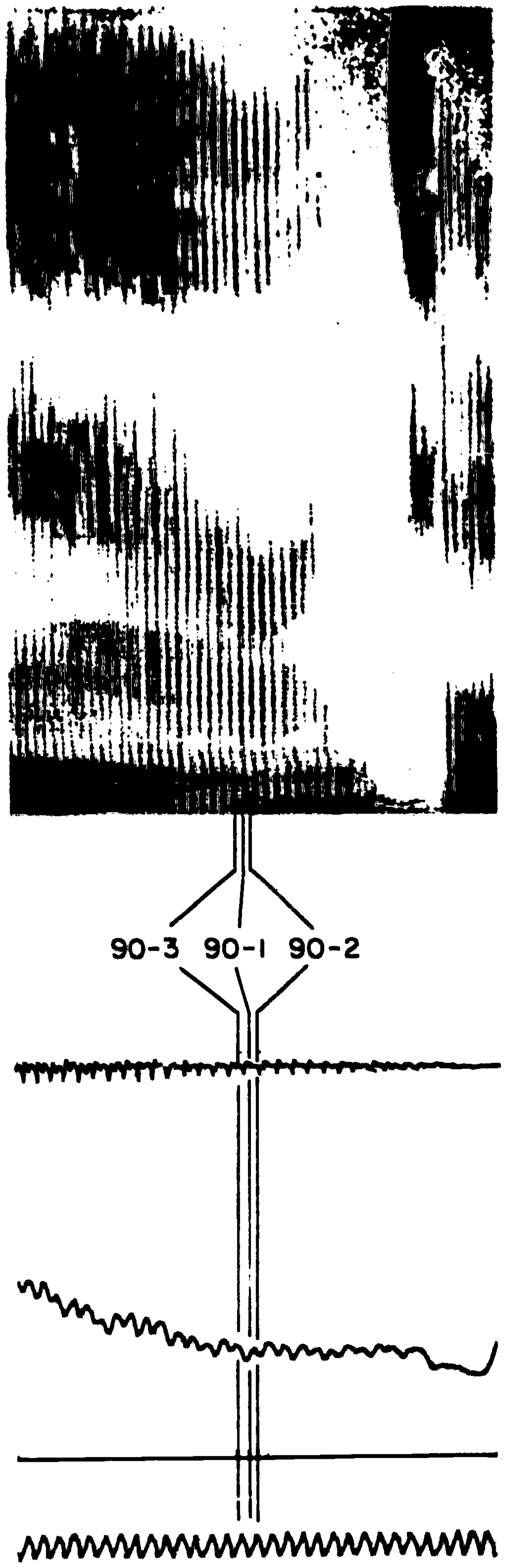

Fig. A57-B

Fig. A57-A,B: Spectrogram and Mingogram Sections for Syllables 89 and 90

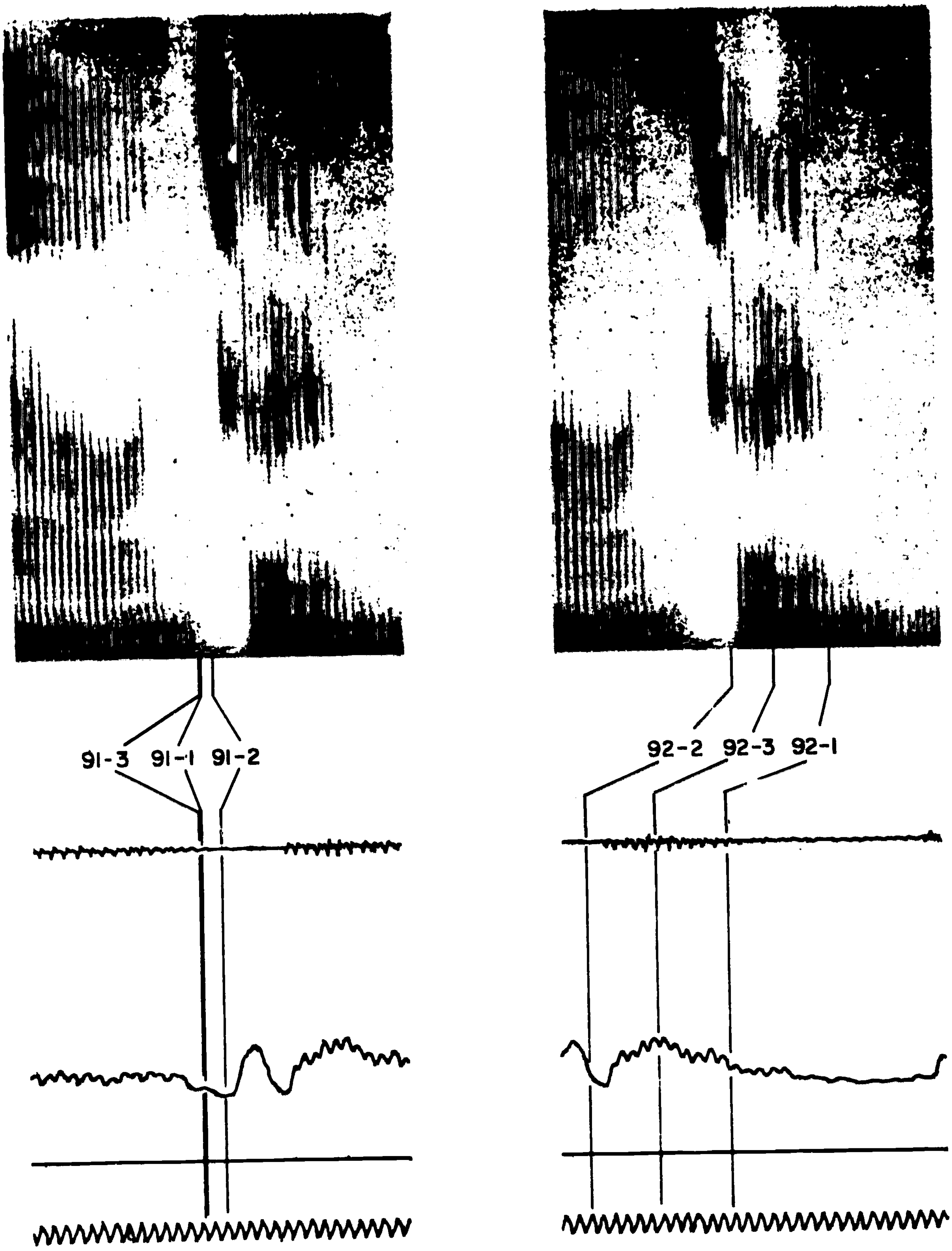

Fig. A58-A

Fig. A58-B

Fig. A58-A,B: Spectrogram and Mingogram Sections for Syllables 91 and 92

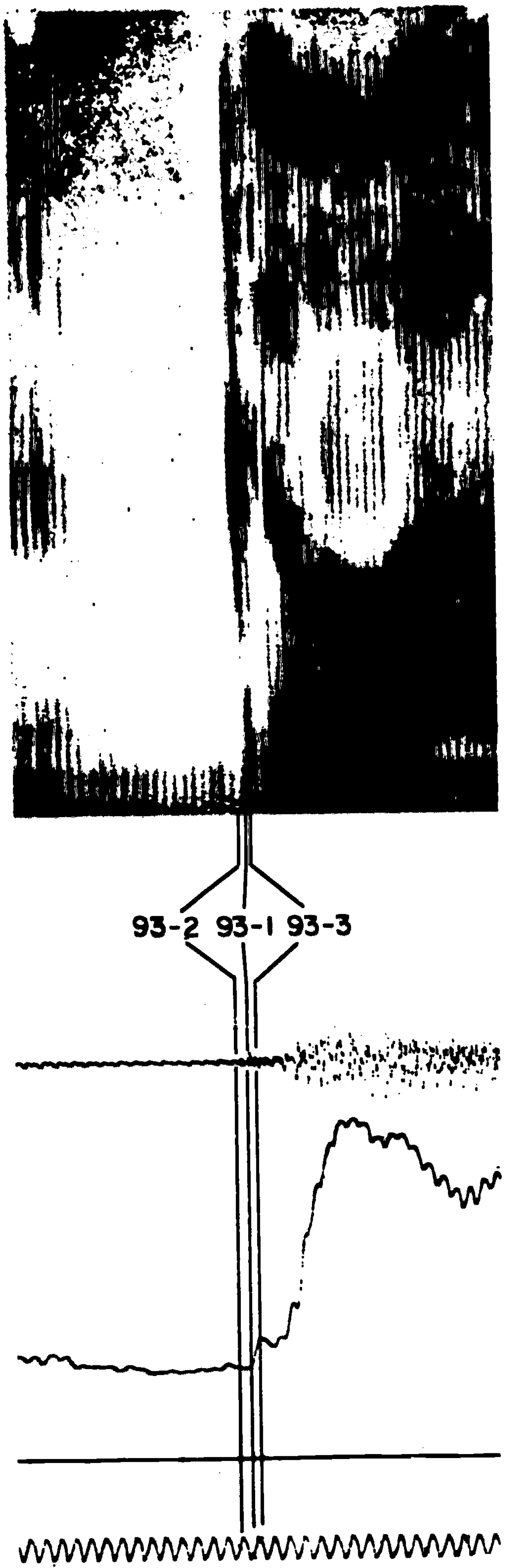

Fig. A59-A

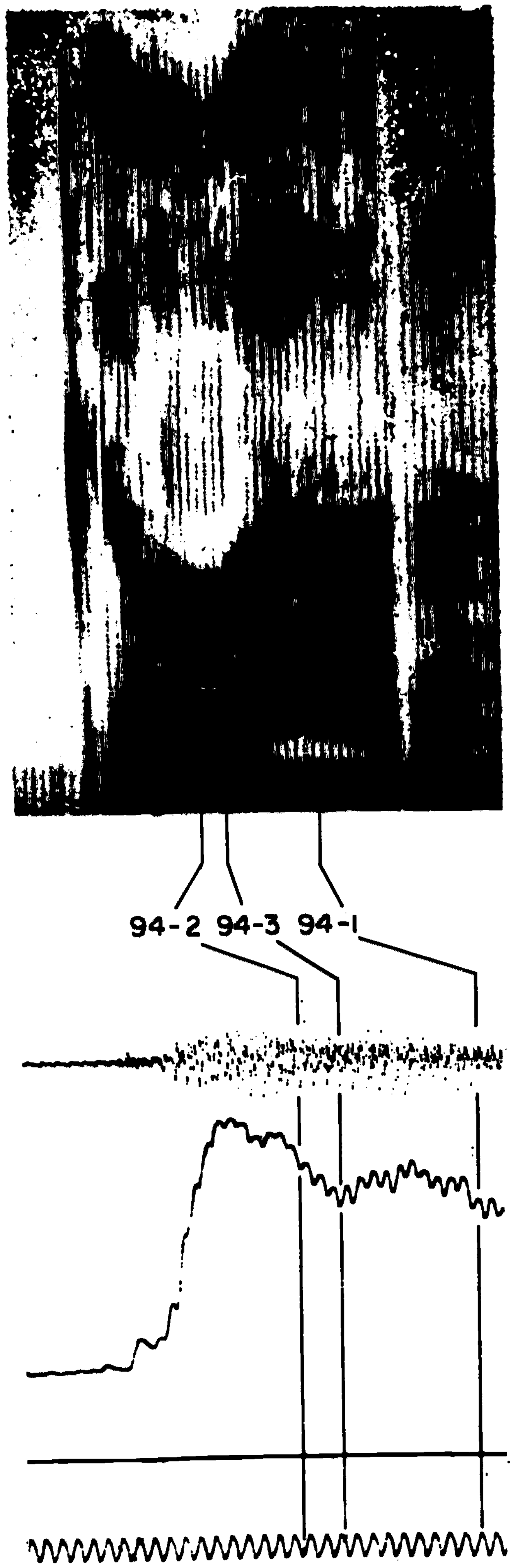

Fig. A59-B

Fig. A59-A,B: Spectrogram and Mingogram Sections for Syllables 93 and 94

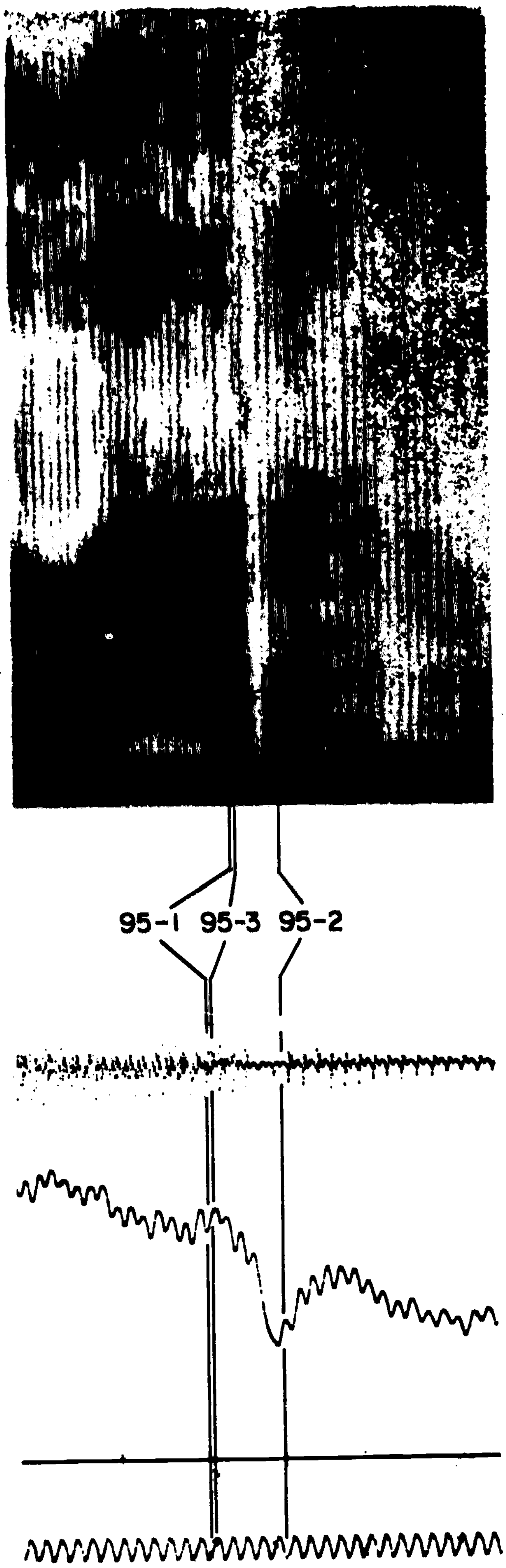

Fig. A60-A

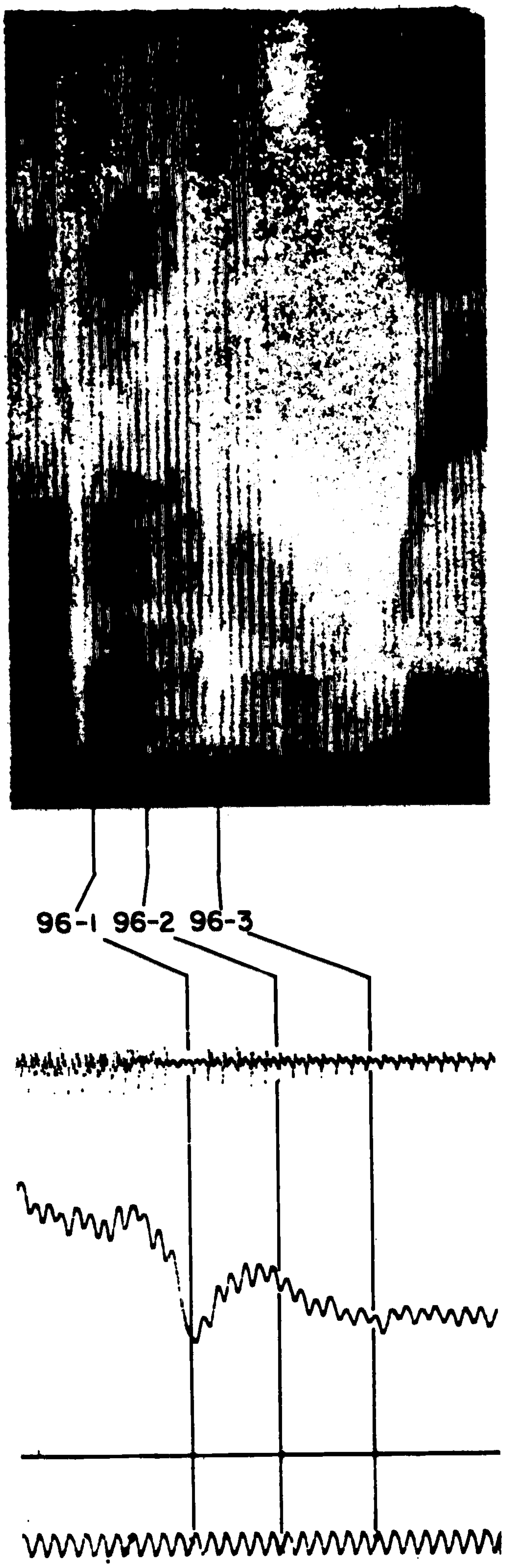

Fig. A60-B

Fig. A60-A,B: Spectrogram and Mingogram Sections for Syllables 95 and 96

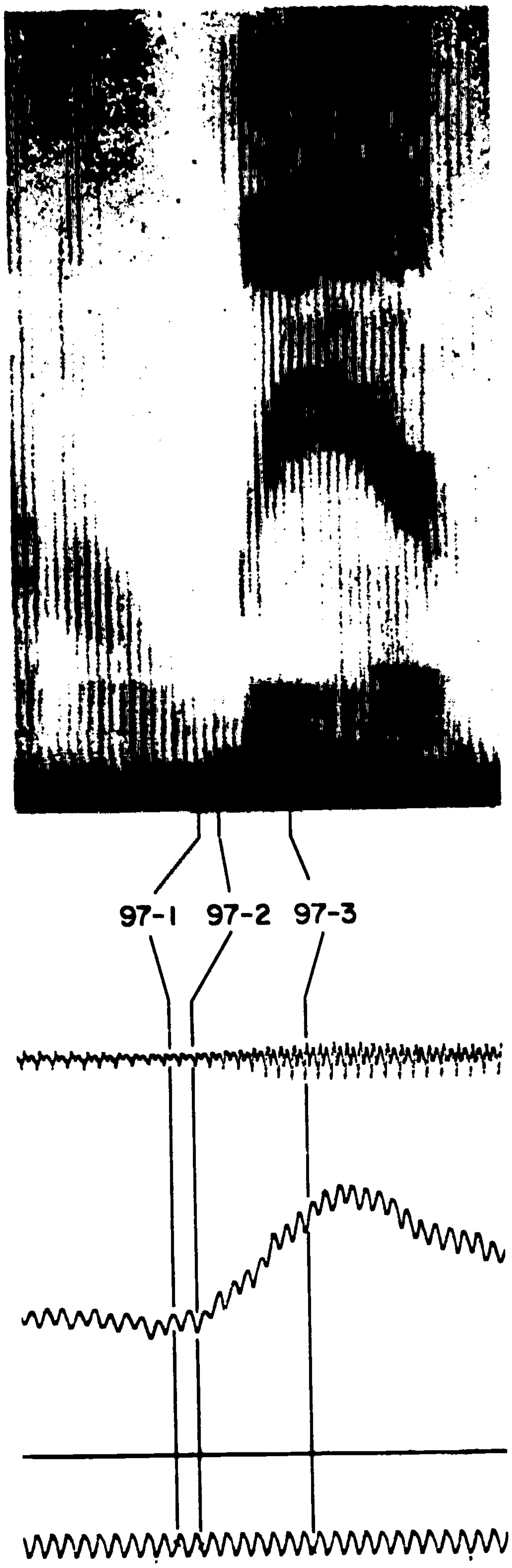

Fig. A61

Fig. A61: Spectrogram and Mingogram
Sections for Syllable 97

APPENDIX B - Click Device

The function of the device illustrated in Figure B1 is to output a pulse at an adjustable time after a pulse arrives at its input. The three components of the device are a Schmitt trigger and two monostable multivibrators. The Schmitt trigger is a threshold circuit which has two output states. If the input signal is greater than +0.5 volts d.c., the output is 0 volts d.c. If the input is less than -0.5 volts d.c., the output is -12 volts d.c. If the input lies between $\pm$0.5 volts d.c., the output stays in its last state. The reason for the inclusion of the Schmitt trigger is that when the timing pulse on the tape loop is input, the output is a sharp, positive going pulse, suitable as input to the first monostable multivibrator.

The first monostable multivibrator produces a negative-going pulse, whose length is adjusted by changing a potentiometer setting. When the positive-going edge of the putput pulse from the Schmitt trigger is input to this device, its output state changes form 0 volts d.c. to -12 volts d.c. The length of time that the -12 volts state is held is a function of the resistance R_2, with greater values of R_2 yielding longer pulses. The shortest possible pulse is 75 msec, and the longest is 1389 msec.

The second monostable multivibrator produces a positive going pulse at the time of the offset of the pulse produced by the first monostable multivibrator. The pulse goes to 0 volts d.c. from -12 volts d.c. and lasts 200 microsec. Various pulse lengths were tried and the 200 microsec value was chosen for its combined sharpness and audibility.

The variability in time between the firing of the Schmitt trigger and the occurrence of the output pulse from the second monostable multivibrator was calibrated and found to be less than 10^{-4} sec at all settings of the potentiometer R_2.

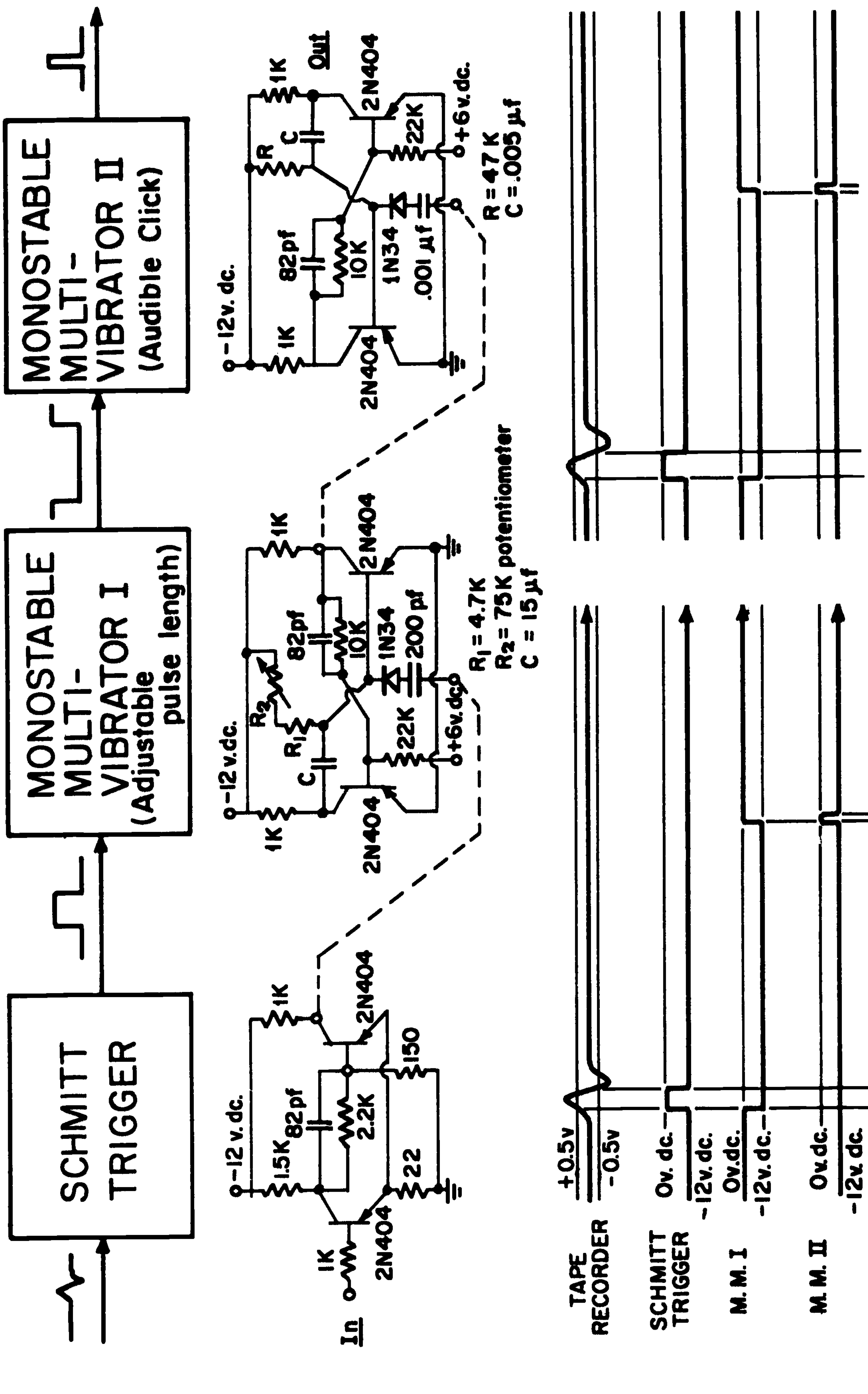

Fig. B1. Click Device

Appendices C and D

pages 172 - 189

available from author on request.

Bibliography

Bolton, T. L. Rhythm. *Amer. J. Psychol.*, 1894, *6*(2).

Classe, A. *The Rhythm of English Prose.* Oxford: Blackwell, 1939.

Cooker, H. S. *Time Relationships of Chest Wall Movements and Intra-Oral Pressures during Speech.* Ph.D. Dissertation, State University of Iowa, 1963, University Microfilms #64-3357.

Draper, M. M., Ladefoged, P., & Whitteridge, D. Expiratory muscles involved in speech. *J. of Physiol.*, 1957, *138*, 17-18.

Draper, M. M. Expiratory pressures and airflow during speech. *Brit. med. J.*, 1960, 1837-1843.

Eblen, R. E., Jr. *A Methodological Study of Electromyography Applied to Speech Breathing.* Ph.D. Dissertation, University of Iowa, 1961.

Fodor, J. A., & Bever, T. G. The Psychological Reality of Linguistic Segments. *J. verb Learn. verb Behav.*, 1965, *4*, 414-420.

Fonagy, I. Elektrophysiologische Beitrage zur Akzentfrage. *Phonetica*, 1958, *2*, 12-58.

Fraisse, P. *The Psychology of Time.* New York: Harper & Row, 1963.

Halliday, M. A. K. The tones of English. *Archivum Linguisticum*, 1963, *15* (1), 1-28.

Hollister, R. D. T. Relation between hand and voice impulse movements. *Speech Monogr.*, 1937, *4* (1), 75-100.

Hoshiko, M. S. Sequence of action of breathing muscles during speech. *J. Speech Hear. Res.*, 1960, *3* (3), 291-297.

Hoshiko, M. S. Electromyographic investigation of the intercostal muscles during speech. *Arch. Phys. Med. & Rehabil.*, March, 1962, 115-119.

Johnson, W. S. Researches in practice and habit. *Stud. from the Yale Psychol. Lab.*, 1898, *6*, 51-103.

Ladefoged, P. The regulation of sub-glottal pressure. *Fol. Phoniat.*, 1960, *12*, 169-175.

Ladefoged, P., & Broadbent, D. E. Perception of Sequence in Auditory Events. *Quat. J. of Exper. Psychol.* 1960, *12*, 162-170.

Ladefoged, P., Draper, M. H., & Whitteridge, D. Syllables and stress. *Miscell. Phonet.*, 1958, *3*, 1-14.

Merrill, W. J., & Bennett, C. A. The application of temporal correlation techniques in psychology. *J. appl. Psychol.*, 1956, *40* (4), 272-280.

Miyake, I. Researches on rhythmic action. *Stud. from the Yale Psychol. Lab.*, 1902, *10*, 1-48.

Moore, P. G. The properties of the mean square successive difference in samples from various populations. *J. Amer. Statist. Assn.*, 1955, *50* (270), 434-456.

Newcomb, W. B. *Some Tempo Manifestations of the Terminals of English.* Ph.D. Dissertation, University of Wisconsin, 1960, University Microfilms #60-1013.

Newcomb, W. B. The perceptual basis of syllable boundary. Unpublished manuscript, General Dynamics Corporation, Rochester, New York, 1961.

Paillard, J. Quelques données psychophysiologiques relatives au déclenchement de la commande motrice. *L'Annee Psychologique*, 1946-47, *47-48*, 28-47.

Peterson, G. E. Some observations on speech. *Quart. J. Speech*, 1958, *44*, 402-412.

Ruckmich, C. A. The rôle of kinaesthesis in the perception of rhythm. *Amer. J. Psychol.*, 1913, *24*, 305-359.

Scheffé, H. *The Analysis of Variance.* New York: Wiley & Sons, 1959.

Scripture, E. W. *The New Psychology*. New York: Scribner, 1897.

Scripture, E. W. Observations on rhythmic action. *Stud. from the Yale Psychol. Lab.*, 1899, *7*, 102-108.

Shen, Y., & Peterson, G. G. Isochronism in English. *Stud. in Ling. Occasional Papers 9*, 1962.

Smith, J. E. K. The precision of the psychophysical method of limits. *Amer. Psychologist*, 1957, *12*, 469. (abstract)

Stetson, R. H. A motor theory of rhythm and discrete succession. *Psychol. Rev.*, 1905, *12*, 250-270, 293-350.

Stetson, R. H. *Motor Phonetics*. Amsterdam: North-Holland, 1951.

Trager, G. L. The theory of accentual syscems. *Language Culture and Personality: Essays in Memory of Edward Sapir*. Wisconsin: Menasha, 1941.

Walker, Helen M. Degrees of freedom. *J. of Educ. Psychol.*, 1940, *31*, 253-260.

Walker, Helen M., & Lev, J. *Statistical Inference*. New York: Holt, Rinehart & Winston, 1953.

Wallin, J. E. W. Researches on the rhythm of speech. *Stud. from the Yale Psychol. Lab.*, 1901, *9*, 1-142.

Winer, B. J. *Statistical Principles in Experimental Design*. New York: McGraw-Hill, 1962.

Woodrow, H. A quantitative study of rhythm. *Arch. of Psychol.*, 1909, (14).

Woodrow, H. Time perception. In S. S. Stevens (Ed.), *Handbook of Expermental Psychology*. New York: Wiley & Sons, 1951. Pp. 1224-1236.

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF FIGURES